aws

Guía para desarrolladores de bases de datos

# Amazon Redshift

Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

# Amazon Redshift: Guía para desarrolladores de bases de datos

Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Las marcas comerciales y la imagen comercial de Amazon no se pueden utilizar en relación con ningún producto o servicio que no sea de Amazon, de ninguna manera que pueda causar confusión entre los clientes y que menosprecie o desacredite a Amazon. Todas las demás marcas registradas que no son propiedad de Amazon son propiedad de sus respectivos propietarios, que pueden o no estar afiliados, conectados o patrocinados por Amazon.

# Table of Contents

Introducción ..... 1
Requisitos previos para utilizar Amazon Redshift ..... 1
Arquitectura de Amazon Redshift ..... 2
Arquitectura del almacenamiento de datos ..... 2
Rendimiento ..... 6
Almacenamiento en columnas ..... 9
Administración de la carga de trabajo ..... 11
Uso de Amazon Redshift con otros servicios ..... 13
Base de datos de muestra ..... 14
Prácticas recomendadas ..... 22
Realización de una prueba de concepto ..... 23
Paso 1: defina el alcance del POC ..... 23
Paso 2: inicie Amazon Redshift ..... 25
Paso 3: cargue los datos ..... 26
Paso 4: analice sus datos ..... 27
Paso 5: optimice ..... 30
Prácticas recomendadas para el diseño de tablas ..... 31
Elección de la clave de clasificación recomendada ..... 31
Elegir el modo de distribución recomendado ..... 32
Uso de la compresión automática ..... 33
Definición de restricciones ..... 34
Uso del menor tamaño posible de columna ..... 35
Uso de tipos de datos de fecha/hora para columnas de fecha ..... 35
Prácticas recomendadas para la carga de datos ..... 35
Obtenga información sobre cómo cargar datos ..... 36
Uso del comando COPY para cargar datos ..... 36
Uso de un único comando COPY ..... 36
Dividir los datos de carga ..... 36
Compresión de los archivos de datos ..... 37
Control de los archivos de datos antes y después de una carga ..... 38
Uso de una inserción de filas múltiples ..... 38
Uso de una inserción masiva ..... 38
Carga de datos en orden de clave de ordenación ..... 39
Carga de datos en bloques secuenciales ..... 39

Uso de tablas de series temporales ... 40
Programación de periodos de mantenimiento ... 40
Prácticas recomendadas para el diseño de consultas ... 41
Seguimiento de las recomendaciones de Advisor ... 43
Regiones de Amazon Redshift en las que se admite Advisor ... 44
Visualización de recomendaciones de Advisor ... 45
Recomendaciones de Advisor ... 46
Tutoriales ... 62
Optimización de tablas automática ... 64
Habilitación, desactivación y supervisión de la optimización de tablas automática ... 65
Habilitación de la optimización automática de tablas ... 65
Eliminación de la optimización automática de tablas de una tabla ... 65
Supervisión de la optimización de tablas automática ... 66
Compresión de columnas ... 67
Codificaciones de compresión ... 68
Prueba de las codificaciones de compresión ... 80
Distribución de datos ... 87
Conceptos de distribución de datos ... 88
Estilos de distribución ... 89
Visualización de los estilos de distribución ... 91
Evaluación de los patrones de consulta ... 92
Designación de estilos de distribución ... 93
Evaluación del plan de consulta ... 94
Ejemplo de plan de consulta ... 97
Ejemplos de distribución ... 101
Claves de clasificación ... 104
Clasificación del diseño de datos multidimensional (versión preliminar) ... 105
Clave de ordenación compuesta ... 107
Clave de ordenación intercalada ... 107
Restricciones de tablas ... 109
Carga de datos ... 110
Carga de tablas con COPY ... 114
Credenciales y permisos de acceso ... 116
Preparación de los datos de entrada ... 118
Carga de datos desde Amazon S3 ... 119
Carga de datos desde Amazon EMR ... 132

Carga de datos desde hosts remotos ... 139
Carga desde Amazon DynamoDB ... 148
Comprobación de carga correcta de datos ... 152
Validación de datos de entrada ... 152
Carga de tablas con compresión automática ... 153
Optimización de tablas angostas ... 156
Carga de valores de columna predeterminados ... 156
Solución de problemas en cargas de datos ... 157
Creación de una integración de eventos de S3 para copiar automáticamente archivos ... 165
Requisitos previos ... 166
Creación de una integración de eventos de S3 ... 171
Creación y supervisión de un COPY JOB ... 173
Consideraciones ... 175
Carga de tablas con DML ... 176
Actualización e inserción ... 176
Realización de una copia profunda ... 185
Análisis de tablas ... 189
Análisis automático ... 189
Análisis de los datos nuevos de las tablas ... 190
Historial del comando ANALYZE ... 195
Limpieza de tablas ... 196
Clasificación automática de tablas ... 197
Eliminación de limpieza automática ... 198
Frecuencia de ejecución de VACUUM ... 198
Fase de ordenación y fase de fusión ... 199
Umbral de limpieza ... 200
Tipos de limpieza ... 200
Cómo minimizar los tiempos de limpieza ... 200
Administración de operaciones de escritura simultáneas ... 209
Aislamiento serializable ... 210
Operaciones de lectura y escritura y de escritura ... 216
Ejemplos de escritura simultánea ... 218
Tutorial: Carga de datos desde Amazon S3 ... 220
Requisitos previos ... 221
Descripción general ... 221
Paso 1: creación de un clúster ... 222

Paso 2: Descargar los archivos de datos ... 223
Paso 3: Cargar los archivos en un bucket de Amazon S3 ... 224
Paso 4: Crear las tablas de muestra ... 226
Paso 5: Ejecutar los comandos COPY ... 228
Paso 6: Limpiar y analizar la base de datos ... 246
Paso 7: Eliminar los recursos ... 247
Resumen ... 247
Descarga de datos ... 249
Descarga de datos en Amazon S3 ... 249
Descarga de archivos de datos cifrados ... 253
Descarga de datos en formato de ancho fijo o delimitado ... 255
Volver a cargar datos descargados ... 256
Funciones definidas por el usuario ... 258
Seguridad y permisos de UDF ... 259
Prevención de conflictos al dar nombre a las UDF ... 259
Sobrecarga de los nombres de función ... 260
Cómo evitar conflictos con las funciones integradas de Amazon Redshift ... 260
UDF de SQL escalares ... 261
Ejemplo ... 261
UDF de Python escalares ... 262
Ejemplo ... 263
Tipos de datos de UDF de Python ... 264
Compatibilidad con el lenguaje Python ... 265
Restricciones ... 270
Registro de errores y advertencias ... 271
UDF de Lambda escalares ... 272
Administración de la seguridad y los permisos de una UDF de Lambda ... 274
Configuración del parámetro de autorización para las UDF de Lambda ... 275
Uso de la interfaz JSON entre Amazon Redshift y Lambda ... 277
Ejemplos de casos de uso para UDF ... 280
Creación de procedimientos almacenados ... 281
Información general de los procedimientos almacenados ... 282
Nomenclatura de los procedimientos almacenados ... 285
Seguridad y privilegios ... 286
Devolución de un conjunto de resultados ... 288
Administración de transacciones ... 290

Errores de captura ... 303
Registro de procedimientos almacenados ... 312
Limitaciones ... 312
Referencia del lenguaje PL/pgSQL ... 314
Convenciones de la referencia de PL/pgSQL ... 314
Estructura de PL/pgSQL ... 315
Instrucciones de PL/pgSQL admitidas ... 321
Vistas materializadas ... 338
Consultas de vistas materializadas ... 341
Reescritura automática de consultas para utilizar vistas materializadas ... 342
Notas de uso ... 342
Cómo funciona la reescritura automática de consultas de vistas materializadas ... 344
Limitaciones ... 344
Vistas materializadas de tablas de lagos de datos externos ... 345
Limitaciones ... 346
Actualización de una vista materializada ... 347
Actualización automática de una vista materializada ... 350
Vistas materializadas automatizadas ... 351
Alcance y consideraciones de SQL de las vistas materializadas automatizadas ... 354
Limitaciones de las vistas materializadas automatizadas ... 354
Facturación para vistas materializadas automatizadas ... 355
Recursos adicionales ... 355
Uso de una función definida por el usuario (UDF) en una vista materializada ... 355
Referencia a una UDF en una vista materializada ... 355
Ingesta de streaming a una vista materializada ... 358
Cómo fluyen los datos de un servicio de streaming a Redshift ... 358
Prácticas recomendadas de análisis de datos para mejorar el rendimiento ... 358
Comportamiento y tipos de datos de ingesta de streaming ... 359
Introducción a la ingesta de streaming de Amazon Kinesis Data Streams ... 365
Introducción a la ingesta de transmisiones desde Apache Kafka ... 370
Tutorial de ingesta de streaming de datos de estaciones de vehículos eléctricos mediante Kinesis ... 385
Vistas del Catálogo de datos ... 389
Requisitos previos ... 390
Ejemplo completo ... 391
Registro seguro ... 392

Consideraciones ... 398
Consulta de datos espaciales ... 400
Tutorial: Uso de funciones SQL espaciales ... 404
Requisitos previos ... 404
Paso 1: Crear tablas y cargar datos de prueba ... 405
Paso 2: Consultar los datos espaciales ... 408
Paso 3: Eliminar los recursos ... 412
Carga de un shapefile ... 412
Terminología ... 414
Cuadro delimitador ... 414
Validez geométrica ... 414
Simplicidad geométrica ... 417
H3 ... 419
Consideraciones ... 419
Consulta de datos con consultas federadas ... 421
Introducción al uso de consultas federadas en PostgreSQL ... 422
Introducción al uso de consultas federadas en PostgreSQL con CloudFormation ... 423
Lanzamiento de una pila de CloudFormation para las consultas federadas de Redshift ... 424
Consulta de datos desde el esquema externo ... 426
Introducción al uso de consultas federadas en MySQL ... 427
Creación de un secreto y rol de IAM ... 428
Requisitos previos ... 428
Ejemplos de uso de una consulta federada ... 430
Ejemplo de uso de una consulta federada con PostgreSQL ... 431
Ejemplo de uso de un nombre compuesto por mayúsculas y minúsculas ... 433
Ejemplo de uso de una consulta federada con MySQL ... 435
Diferencias de tipos de datos ... 436
Consideraciones ... 440
Versiones compatibles de bases de datos federadas ... 442
Amazon Redshift Spectrum ... 444
Información general sobre Amazon Redshift Spectrum ... 444
Regiones de Amazon Redshift Spectrum ... 446
Limitaciones de Amazon Redshift Spectrum ... 446
Introducción a Amazon Redshift Spectrum ... 447
Requisitos previos ... 447
CloudFormation ... 448

Introducción paso a paso a Redshift Spectrum ... 448
Paso 1. Creación de un rol de IAM ... 449
Paso 2: Asociar el rol de IAM a su clúster ... 453
Paso 3: Crear un esquema externo y una tabla externa ... 453
Paso 4: Consultar sus datos en Amazon S3 ... 455
Lanzamiento de su pila de CloudFormation y, a continuación, consulta de sus datos ... 458
Políticas de IAM para Amazon Redshift Spectrum ... 462
Permisos de Amazon S3 ... 463
Permisos entre cuentas de Amazon S3 ... 464
Concesión o restricción del acceso a través de Redshift Spectrum ... 465
Permisos mínimos ... 465
Encadenamiento de roles de IAM ... 467
Acceso a datos de AWS Glue ... 467
Redshift Spectrum y Lake Formation ... 476
Uso de filtros de datos para seguridad en el nivel de fila y de celda ... 478
Archivos de datos para consultas en Amazon Redshift Spectrum ... 478
Formatos de datos para Redshift Spectrum ... 479
Tipos de compresión para Redshift Spectrum ... 480
Cifrado para Redshift Spectrum ... 481
Esquemas externos ... 482
Uso de catálogos externos ... 484
tablas externas ... 489
Pseudocolumnas ... 491
Partición de tablas externas de Redshift Spectrum ... 492
Asignación a columnas ORC ... 498
Creación de tablas externas para datos administrados por Hudi ... 501
Creación de tablas externas para datos de Delta Lake ... 503
Uso de tablas de Apache Iceberg ... 505
Consideraciones al utilizar tablas de Apache Iceberg ... 506
Tipos de datos compatibles ... 508
Rendimiento de las consultas de Amazon Redshift Spectrum ... 510
Opciones de gestión de datos ... 514
Realización de subconsultas correlacionadas ... 514
Métricas ... 515
Solución de problemas de consultas ... 516
Cantidad de reintentos superada ... 517

Acceso limitado ... 517
Límite de recursos superado ... 519
No se han devuelto filas para una tabla particionada ... 519
Error de no autorización ... 519
Formatos de datos incompatibles ... 520
Error de sintaxis al usar el DDL de Hive en Amazon Redshift ... 520
Permiso para crear tablas temporales ... 521
Rango no válido ... 521
Número de versión de Parquet no válido ... 521
Tutorial: Consulta de datos anidados con Amazon Redshift Spectrum ... 522
Descripción general ... 522
Paso 1: Crear una tabla externa que contenga datos anidados ... 523
Paso 2: Consultar los datos anidados en Amazon S3 con extensiones SQL ... 524
Casos de uso de datos anidados ... 528
Limitaciones de los datos anidados (versión preliminar) ... 531
Serialización de datos JSON anidados y complejos ... 533
Bocetos de HyperLogLog ... 537
Consideraciones ... 538
Limitaciones ... 539
Ejemplos ... 539
Ejemplo: devolver la cardinalidad en una subconsulta ... 540
Ejemplo: devolver un tipo HLLSKETCH a partir de bocetos combinados en una subconsulta ... 540
Ejemplo: devolver un boceto de HyperLogLog a partir de la combinación de varios bocetos ... 541
Ejemplo: generar bocetos de HyperLogLog a partir de datos de S3 mediante tablas externas ... 542
Consultas entre bases de datos ... 545
Consideraciones ... 547
Limitaciones ... 548
Ejemplos ... 549
Uso de consultas entre bases de datos con el editor de consultas ... 554
Compatibilidad con Apache Iceberg ... 557
Regiones en las que está disponible la compatibilidad con Apache Iceberg ... 557
Consideraciones y limitaciones al usar catálogos de Amazon Redshift en el AWS Glue Data Catalog ... 558

Requisitos de las políticas de IAM ..... 559
Registro de clústeres y espacios de nombres ..... 560
Anulación del registro de clústeres y espacios de nombres ..... 561
Grupos de trabajo administrados ..... 563
Consulta de catálogos registrados en el AWS Glue Data Catalog ..... 563
Uso compartido de datos ..... 565
Casos de uso compartido de datos ..... 565
Casos de uso de escrituras en varios almacenamientos ..... 566
Uso compartido de datos en niveles diferentes ..... 566
Administración de la coherencia del uso compartido de datos ..... 567
Consideraciones ..... 567
..... 567
Consideraciones generales sobre el uso compartido de datos ..... 568
Consideraciones sobre lecturas y escrituras mediante uso compartido de datos ..... 569
Consideraciones sobre el uso compartido de datos con tablas de lago de datos ..... 571
Consideraciones sobre el uso compartido de datos con AWS Lake Formation ..... 571
Consideraciones sobre el uso compartido de datos con AWS Data Exchange ..... 572
Permisos que puede conceder a los recursos compartidos de datos ..... 574
Instrucciones SQL compatibles con escrituras mediante uso compartido de datos en consumidores ..... 575
Instrucciones SQL no compatibles con escrituras mediante uso compartido de datos en consumidores ..... 576
Regiones de AWS disponibles ..... 577
Introducción ..... 580
Introducción al uso compartido de datos de solo lectura en la consola ..... 581
Introducción al uso compartido de datos de solo lectura con la interfaz de SQL ..... 601
Introducción a las escrituras en varios almacenamientos ..... 649
Introducción a CloudFormation ..... 685
Tipos de recursos compartidos de datos ..... 691
Recursos compartidos de datos estándar ..... 691
Recursos compartidos de datos de AWS Data Exchange ..... 693
Recursos compartidos de datos administrados por AWS Lake Formation ..... 696
Estado del recurso compartido de datos ..... 698
Administración del acceso a las operaciones de la API de uso compartido de datos con las políticas de IAM ..... 700
Conexión a bases de datos de consumidores ..... 702

Acceso a datos compartidos ... 702
Acceso a los metadatos para objetos compartidos ... 702
Integración del uso compartido de datos de Amazon Redshift a las herramientas de inteligencia empresarial ... 703
Monitoreo y auditoría del uso compartido de datos ... 704
Integración del uso compartido de datos de Amazon Redshift con AWS CloudTrail ... 706
Datos semiestructurados en Amazon Redshift ... 707
Casos de uso del tipo de datos SUPER ... 707
Conceptos para el uso del tipo de datos SUPER ... 709
Consideraciones sobre los datos SUPER ... 711
Conjunto de datos de muestra SUPER ... 711
Carga de datos semiestructurados en Amazon Redshift ... 714
Análisis de documentos JSON en columnas SUPER ... 714
Uso de COPY para cargar datos JSON en Amazon Redshift ... 715
Descarga de datos semiestructurados ... 720
Consulta de datos semiestructurados ... 721
Navegación ... 722
Consultas de eliminación del anidamiento ... 723
Anulación de la dinamización de los objetos ... 725
Escritura dinámica ... 726
Semántica laxa ... 730
Tipos de introspección ... 730
Ordenar por ... 732
Operadores y funciones ... 733
Operadores aritméticos ... 733
Funciones aritméticas ... 734
Funciones de matriz ... 734
Configuraciones SUPER ... 736
Modos laxo y estricto de SUPER ... 737
Acceso a campos JSON con letras mayúsculas y con mayúsculas y minúsculas combinadas ... 737
Opciones de análisis ... 739
Limitaciones ... 740
Tipo de datos SUPER y vistas materializadas ... 743
Aceleración de consultas PartiQL ... 743
Limitaciones de uso del tipo de datos SUPER con vistas materializadas ... 747

Machine learning ... 749
Información general sobre machine learning ... 751
Solución de problemas con machine learning ... 752
Términos y conceptos de Amazon Redshift ML ... 754
Machine learning para principiantes y expertos ... 755
Costos de utilizar Amazon Redshift ML ... 758
Costos por el uso de Amazon Redshift ML con SageMaker AI ... 758
Costos por el uso de Amazon Redshift ML con Amazon Bedrock ... 758
Introducción: Amazon Redshift ML ... 760
Configuración administrativa ... 761
Uso de la explicabilidad del modelo con Amazon Redshift ML ... 767
Métricas de probabilidad de Amazon Redshift ML ... 768
Tutoriales de Amazon Redshift ML ... 770
Tutorial: Cree modelos de bajas de clientes ... 772
Tutorial: Creación de modelos de agrupación en clústeres K-means ... 781
Tutorial: Creación de modelos de clasificación multiclase ... 790
Tutorial: Creación de modelos XGBoost ... 800
Tutorial: Creación de modelos de regresión ... 806
Tutorial: Creación de modelos de regresión con aprendizaje lineal ... 820
Tutorial: Creación de modelos de clasificación multiclase con aprendizaje lineal ... 828
Integración de Amazon Redshift ML con Amazon Bedrock ... 850
Creación de un rol de IAM ... 851
Creación de un modelo externo ... 852
Uso de un modelo externo ... 853
Ingeniería de peticiones ... 855
Ajuste del rendimiento de las consultas ... 857
Procesamiento de consultas ... 857
Flujo de trabajo de planificación y ejecución de consultas ... 858
Creación e interpretación de un plan de consultas ... 860
Revisión de los pasos del plan de consulta ... 869
Factores que afectan al rendimiento de las consultas ... 871
Análisis y mejora de consultas ... 873
Flujo de trabajo de análisis de consultas ... 873
Revisión de alertas de consultas ... 875
Análisis del plan de consulta ... 877
Análisis del resumen de consultas ... 878

Mejora de las consultas ... 886
Consultas de diagnóstico para el ajuste de consultas ... 890
Solución de problemas de consultas ... 894
Error en la conexión ... 895
Falta de respuesta de la consulta ... 896
Tiempo de ejecución de consulta demasiado largo ... 897
Error en la carga ... 899
Tiempo de carga excesivo ... 899
Datos de carga incorrectos ... 900
Configuración del parámetro de tamaño de búsqueda de la JDBC ... 900
Administración de la carga de trabajo ... 902
Cambio del modo de WLM ... 904
Modificación de la configuración de WLM ... 904
Migración desde WLM manual a WLM automática ... 904
WLM automática ... 906
Priority (Prioridad) ... 908
Modo de escalado de simultaneidad ... 908
Grupos de usuarios ... 908
Roles de usuario ... 908
Grupos de consultas ... 908
Caracteres comodín ... 909
Reglas de monitorización de consultas ... 909
Comprobación de WLM automática ... 909
Prioridad de consulta ... 910
WLM manual ... 915
Modo de escalado de simultaneidad ... 917
Nivel de simultaneidad ... 917
Grupos de usuarios ... 920
Roles de usuario ... 920
Grupos de consultas ... 920
Caracteres comodín ... 920
Porcentaje de memoria de WLM por utilizar ... 921
Tiempo de espera de WLM ... 921
Reglas de monitorización de consultas ... 922
Salto de cola de consultas de WLM ... 922
Tutorial: Configuración de colas de WLM manual ... 926

Escalado de simultaneidad ... 945
Capacidades de escalado de simultaneidad ... 946
Limitaciones para el escalado de simultaneidad ... 946
Regiones para el escalado de simultaneidad ... 948
Candidatos al escalado de simultaneidad ... 949
Configuración de colas de escalado de simultaneidad ... 912
Monitoreo del escalado de simultaneidad ... 950
Vistas del sistema de escalado de simultaneidad ... 951
Aceleración de consultas cortas ... 952
Tiempo máximo de ejecución de SQA ... 953
Monitorización de SQA ... 953
Reglas de asignación de colas de WLM ... 954
Ejemplo de asignaciones de colas ... 957
Asignación de consultas a las colas ... 959
Asignación de consultas a las colas en función de los roles de usuarios ... 959
Asignación de consultas a las colas en función de los grupos de usuarios ... 960
Asignación de una consulta a un grupo de consultas ... 960
Asignación de consultas a la cola de superusuario ... 961
Propiedades dinámicas y estáticas ... 962
Asignación de memoria dinámica de WLM ... 963
Ejemplo de WLM dinámica ... 964
Reglas de monitorización de consultas ... 967
Definición de una regla de monitoreo de consultas ... 968
Métricas de supervisión de consultas para Amazon Redshift aprovisionado ... 970
Métricas de supervisión de consultas para Amazon Redshift sin servidor ... 974
Plantillas de reglas de monitoreo de consultas ... 975
Tablas y vistas de sistema para reglas de monitoreo de consultas ... 977
Tablas y vistas de sistema de WLM ... 978
ID de la clase de servicio WLM ... 980
Seguridad de base de datos ... 981
Información general de la seguridad de Amazon Redshift ... 982
Permisos de usuario de base de datos predeterminados ... 983
Superusuarios ... 984
Usuarios ... 985
Creación, modificación y eliminación de usuarios ... 985
Grupos ... 986

Creación, modificación y eliminación de grupos ... 987
Ejemplo del control de acceso de usuarios y grupos ... 987
Esquemas ... 989
Ruta de búsqueda ... 989
Creación, modificación y eliminación de esquemas ... 990
Permisos ... 990
Control de acceso con base en roles ... 991
Jerarquía de roles ... 992
Asignación de roles ... 992
Roles definidos por el sistema de Amazon Redshift ... 993
Permisos del sistema ... 997
Permisos de objetos de bases de datos ... 1010
ALTER DEFAULT PRIVILEGES para RBAC ... 1010
Consideraciones sobre el uso de roles ... 1010
Administración de roles ... 1011
Tutorial: creación de roles y realización de consultas con RBAC ... 1011
Seguridad de nivel básico ... 1031
Uso de políticas de RLS en instrucciones SQL ... 1032
Combinación de varias políticas por usuario ... 1032
Propiedad y administración de la política de RLS ... 1034
Objetos y principios dependientes de políticas ... 1036
Consideraciones y limitaciones ... 1038
Prácticas recomendadas ... 1042
Ejemplo completo ... 1044
Seguridad de los metadatos ... 1048
Enmascaramiento de datos dinámico ... 1050
Comandos SQL para políticas de DDM ... 1050
Jerarquía de políticas de DDM ... 1052
Uso del DDM con rutas de tipo SUPER ... 1054
Enmascaramiento dinámico y condicional de datos ... 1059
Vistas del sistema de DDM ... 1060
Consideraciones ... 1062
Ejemplo completo ... 1067
Permisos acotados ... 1071
Consideraciones ... 1071
Referencia de SQL ... 1073

SQL en Amazon Redshift ..... 1073
Funciones SQL admitidas en el nodo principal ..... 1074
Amazon Redshift y PostgreSQL ..... 1076
Uso de SQL ..... 1085
Convenciones de referencia a SQL ..... 1085
Elementos básicos ..... 1086
Expressions ..... 1143
Condiciones ..... 1148
Comandos SQL ..... 1177
ABORT ..... 1182
ALTER DATABASE ..... 1183
ALTER DATASHARE ..... 1189
ALTER DEFAULT PRIVILEGES ..... 1193
ALTER EXTERNAL SCHEMA ..... 1197
ALTER EXTERNAL VIEW ..... 1199
ALTER FUNCTION ..... 1201
ALTER GROUP ..... 1202
MODIFICAR PROVEEDOR DE IDENTIDADES ..... 1203
ALTER MASKING POLICY ..... 1206
ALTER MATERIALIZED VIEW ..... 1207
ALTER RLS POLICY ..... 1215
ALTER ROLE ..... 1216
ALTER PROCEDURE ..... 1217
ALTER SCHEMA ..... 1219
ALTER SYSTEM ..... 1221
ALTER TABLE ..... 1223
ALTER TABLE APPEND ..... 1250
ALTER USER ..... 1256
ANALYZE ..... 1262
ANALYZE COMPRESSION ..... 1266
ATTACH MASKING POLICY ..... 1268
ATTACH RLS POLICY ..... 1270
BEGIN ..... 1272
CALL ..... 1274
CANCEL ..... 1278
CLOSE ..... 1280

COMMENT ..... 1281
COMMIT ..... 1284
COPY ..... 1285
CREATE DATABASE ..... 1393
CREATE DATASHARE ..... 1413
CREATE EXTERNAL FUNCTION ..... 1414
CREATE EXTERNAL MODEL ..... 1426
CREATE EXTERNAL SCHEMA ..... 1430
CREATE EXTERNAL TABLE ..... 1442
CREATE EXTERNAL VIEW ..... 1473
CREATE FUNCTION ..... 1474
CREATE GROUP ..... 1481
CREATE IDENTITY PROVIDER ..... 1482
CREATE LIBRARY ..... 1485
CREATE MASKING POLICY ..... 1488
CREATE MATERIALIZED VIEW ..... 1489
CREATE MODEL ..... 1495
CREATE PROCEDURE ..... 1527
CREATE RLS POLICY ..... 1534
CREAR ROL ..... 1536
CREATE SCHEMA ..... 1537
CREATE TABLE ..... 1540
CREATE TABLE AS ..... 1565
CREAR USUARIO ..... 1578
CREATE VIEW ..... 1586
DEALLOCATE ..... 1592
DECLARE ..... 1593
DELETE ..... 1597
DESC DATASHARE ..... 1601
DESC IDENTITY PROVIDER ..... 1602
DETACH MASKING POLICY ..... 1603
DETACH RLS POLICY ..... 1604
DROP DATABASE ..... 1605
DROP DATASHARE ..... 1607
DROP EXTERNAL VIEW ..... 1608
DROP FUNCTION ..... 1609

DROP GROUP ..... 1611
DROP IDENTITY PROVIDER ..... 1612
DROP LIBRARY ..... 1613
DROP MASKING POLICY ..... 1614
DROP MODEL ..... 1614
DROP MATERIALIZED VIEW ..... 1615
DROP PROCEDURE ..... 1616
DROP RLS POLICY ..... 1618
DROP ROLE ..... 1618
DROP SCHEMA ..... 1620
DROP TABLE ..... 1622
DROP USER ..... 1626
DROP VIEW ..... 1628
END ..... 1631
EXECUTE ..... 1632
EXPLAIN ..... 1633
FETCH ..... 1642
GRANT ..... 1644
INSERT ..... 1671
INSERT (tabla externa) ..... 1679
LOCK ..... 1682
MERGE ..... 1683
PREPARE ..... 1690
REFRESH MATERIALIZED VIEW ..... 1692
RESET ..... 1696
REVOKE ..... 1697
ROLLBACK ..... 1717
SELECT ..... 1719
SELECT INTO ..... 1797
SET ..... 1798
SET SESSION AUTHORIZATION ..... 1804
SET SESSION CHARACTERISTICS ..... 1805
SHOW ..... 1805
SHOW COLUMNS ..... 1806
SHOW EXTERNAL TABLE ..... 1809
SHOW DATABASES ..... 1812

SHOW GRANTS ... 1816
SHOW MODEL ... 1822
SHOW DATASHARES ... 1825
SHOW PROCEDURE ... 1826
MOSTRAR ESQUEMAS ... 1828
SHOW TABLE ... 1830
SHOW TABLES ... 1832
SHOW VIEW ... 1834
START TRANSACTION ... 1835
TRUNCATE ... 1835
UNLOAD ... 1837
UPDATE ... 1873
USE ... 1881
VACUUM ... 1891
Referencia de funciones SQL ... 1899
Funciones específicas del nodo principal ... 1900
Funciones de agregación ... 1902
Funciones de matriz ... 1932
Funciones de agregación bit a bit ... 1937
Expresiones condicionales ... 1945
Funciones de formato de tipo de datos ... 1960
Funciones de fecha y hora ... 2000
Funciones hash ... 2074
Funciones HyperLogLog ... 2085
Funciones JSON ... 2090
Funciones de machine learning. ... 2112
Funciones matemáticas ... 2116
Funciones de objeto ... 2156
Funciones espaciales ... 2166
Funciones de cadena ... 2312
Funciones de información acerca del tipo SUPER ... 2394
Funciones VARBYTE ... 2409
Funciones de ventana ... 2419
Funciones de administración del sistema ... 2488
Funciones de información del sistema ... 2499
Palabras reservadas ... 2533

Referencia de las tablas y vistas de sistema ..... 2537
Tipos de tablas y vistas de sistema ..... 2538
Visibilidad de datos en las tablas y vistas de sistema ..... 2539
Filtrado de consultas generadas por el sistema ..... 2540
Migración de consultas aprovisionadas solo a consultas de vista de supervisión de SYS ..... 2540
Migración de clústeres aprovisionados a Amazon Redshift sin servidor ..... 2540
Actualización de consultas mientras se permanece en un clúster aprovisionado ..... 2541
Mejora del seguimiento de los identificadores de consultas mediante las vistas de monitoreo SYS ..... 2541
Ejemplo ..... 2542
Identificadores de consulta, proceso y sesión de tablas del sistema ..... 2549
Vistas de metadatos SVV ..... 2549
SVV_ACTIVE_CURSORS ..... 2552
SVV_ALL_COLUMNS ..... 2553
SVV_ALL_SCHEMAS ..... 2555
SVV_ALL_TABLES ..... 2557
SVV_ALTER_TABLE_RECOMMENDATIONS ..... 2558
SVV_ATTACHED_MASKING_POLICY ..... 2560
SVV_COLUMNS ..... 2563
SVV_COLUMN_PRIVILEGES ..... 2565
SVV_COPY_JOB_INTEGRATIONS ..... 2567
SVV_DATABASE_PRIVILEGES ..... 2568
SVV_DATASHARE_PRIVILEGES ..... 2570
SVV_DATASHARES ..... 2571
SVV_DATASHARE_CONSUMERS ..... 2574
SVV_DATASHARE_OBJECTS ..... 2575
SVV_DEFAULT_PRIVILEGES ..... 2578
SVV_DISKUSAGE ..... 2579
SVV_EXTERNAL_COLUMNS ..... 2582
SVV_EXTERNAL_DATABASES ..... 2584
SVV_EXTERNAL_PARTITIONS ..... 2584
SVV_EXTERNAL_SCHEMAS ..... 2585
SVV_EXTERNAL_TABLES ..... 2587
SVV_FUNCTION_PRIVILEGES ..... 2588
SVV_GEOGRAPHY_COLUMNS ..... 2590
SVV_GEOMETRY_COLUMNS ..... 2591

SVV_IAM_PRIVILEGES ..... 2593
SVV_IDENTITY_PROVIDERS ..... 2594
SVV_INTEGRATION ..... 2596
SVV_INTEGRATION_TABLE_MAPPING ..... 2598
SVV_INTEGRATION_TABLE_STATE ..... 2601
SVV_INTERLEAVED_COLUMNS ..... 2603
SVV_LANGUAGE_PRIVILEGES ..... 2604
SVV_MASKING_POLICY ..... 2605
SVV_ML_MODEL_INFO ..... 2606
SVV_ML_MODEL_PRIVILEGES ..... 2607
SVV_MV_DEPENDENCY ..... 2609
SVV_MV_INFO ..... 2610
SVV_QUERY_INFLIGHT ..... 2613
SVV_QUERY_STATE ..... 2614
SVV_REDSHIFT_COLUMNS ..... 2618
SVV_REDSHIFT_DATABASES ..... 2621
SVV_REDSHIFT_FUNCTIONS ..... 2623
SVV_REDSHIFT_SCHEMA_QUOTA ..... 2624
SVV_REDSHIFT_SCHEMAS ..... 2625
SVV_REDSHIFT_TABLES ..... 2627
SVV_RELATION_PRIVILEGES ..... 2628
SVV_RLS_APPLIED_POLICY ..... 2630
SVV_RLS_ATTACHED_POLICY ..... 2632
SVV_RLS_POLICY ..... 2633
SVV_RLS_RELATION ..... 2634
SVV_ROLE_GRANTS ..... 2636
SVV_ROLES ..... 2637
SVV_SCHEMA_PRIVILEGES ..... 2638
SVV_SCHEMA_QUOTA_STATE ..... 2640
SVV_SYSTEM_PRIVILEGES ..... 2641
SVV_TABLE_INFO ..... 2642
SVV_TABLES ..... 2647
SVV_TRANSACTIONS ..... 2648
SVV_USER_GRANTS ..... 2650
SVV_USER_INFO ..... 2652
SVV_VACUUM_PROGRESS ..... 2653

SVV_VACUUM_SUMMARY ..... 2655
Vistas de monitoreo de SYS ..... 2658
SYS_ANALYZE_COMPRESSION_HISTORY ..... 2659
SYS_ANALYZE_HISTORY ..... 2662
SYS_APPLIED_MASKING_POLICY_LOG ..... 2664
SYS_AUTO_TABLE_OPTIMIZATION ..... 2666
SYS_CHILD_QUERY_TEXT ..... 2668
SYS_CONNECTION_LOG ..... 2669
SYS_COPY_JOB ..... 2673
SYS_COPY_JOB_DETAIL ..... 2674
SYS_COPY_JOB_INFO ..... 2676
SYS_COPY_REPLACEMENTS ..... 2677
SYS_DATASHARE_CHANGE_LOG ..... 2678
SYS_DATASHARE_CROSS_REGION_USAGE ..... 2682
SYS_DATASHARE_USAGE_CONSUMER ..... 2683
SYS_DATASHARE_USAGE_PRODUCER ..... 2684
SYS_EXTERNAL_QUERY_DETAIL ..... 2686
SYS_EXTERNAL_QUERY_ERROR ..... 2689
SYS_INTEGRATION_ACTIVITY ..... 2692
SYS_INTEGRATION_TABLE_ACTIVITY ..... 2694
SYS_INTEGRATION_TABLE_STATE_CHANGE ..... 2695
SYS_LOAD_DETAIL ..... 2697
SYS_LOAD_ERROR_DETAIL ..... 2700
SYS_LOAD_HISTORY ..... 2703
SYS_MV_REFRESH_HISTORY ..... 2707
SYS_MV_STATE ..... 2710
SYS_PROCEDURE_CALL ..... 2713
SYS_PROCEDURE_MESSAGES ..... 2715
SYS_QUERY_DETAIL ..... 2717
SYS_QUERY_EXPLAIN ..... 2726
SYS_QUERY_HISTORY ..... 2730
SYS_QUERY_TEXT ..... 2743
SYS_RESTORE_LOG ..... 2746
SYS_RESTORE_STATE ..... 2749
SYS_SCHEMA_QUOTA_VIOLATIONS ..... 2751
SYS_SERVERLESS_USAGE ..... 2752

SYS_SESSION_HISTORY ... 2756
SYS_SPATIAL_SIMPLIFY ... 2757
SYS_STREAM_SCAN_ERRORS ... 2759
SYS_STREAM_SCAN_STATES ... 2760
SYS_TRANSACTION_HISTORY ... 2762
SYS_UDF_LOG ... 2765
SYS_UNLOAD_DETAIL ... 2767
SYS_UNLOAD_HISTORY ... 2769
SYS_USERLOG ... 2771
SYS_VACUUM_HISTORY ... 2773
Asignación de la vista de sistema para migrar a las vistas de supervisión SYS ... 2777
SYS_QUERY_HISTORY ... 2778
SYS_QUERY_DETAIL ... 2779
SYS_RESTORE_LOG ... 2780
SYS_RESTORE_STATE ... 2780
SYS_TRANSACTION_HISTORY ... 2781
SYS_QUERY_TEXT ... 2781
SYS_CONNECTION_LOG ... 2781
SYS_SESSION_HISTORY ... 2781
SYS_LOAD_DETAIL ... 2781
SYS_LOAD_HISTORY ... 2782
SYS_LOAD_ERROR_DETAIL ... 2782
SYS_UNLOAD_HISTORY ... 2782
SYS_UNLOAD_DETAIL ... 2782
SYS_COPY_REPLACEMENTS ... 2782
SYS_DATASHARE_USAGE_CONSUMER ... 2783
SYS_DATASHARE_USAGE_PRODUCER ... 2783
SYS_DATASHARE_CROSS_REGION_USAGE ... 2783
SYS_DATASHARE_CHANGE_LOG ... 2783
SYS_EXTERNAL_QUERY_DETAIL ... 2783
SYS_EXTERNAL_QUERY_ERROR ... 2784
SYS_VACUUM_HISTORY ... 2784
SYS_ANALYZE_HISTORY ... 2784
SYS_ANALYZE_COMPRESSION_HISTORY ... 2784
SYS_MV_REFRESH_HISTORY ... 2784
SYS_MV_STATE ... 2785

SYS_PROCEDURE_CALL ... 2785
SYS_PROCEDURE_MESSAGES ... 2785
SYS_UDF_LOG ... 2785
SYS_USERLOG ... 2785
SYS_SCHEMA_QUOTA_VIOLATIONS ... 2785
SYS_SPATIAL_SIMPLIFY ... 2786
Monitoreo del sistema (solo aprovisionado) ... 2786
Vistas STL para registro ... 2786
Tablas STV para datos de instantáneas ... 2934
Vistas de SVCS para clústeres de escalado principales y de simultaneidad ... 2995
Vistas SVL para el clúster principal ... 3024
Tablas de catálogos de sistema ... 3103
PG_ATTRIBUTE_INFO ... 3103
PG_CLASS_INFO ... 3104
PG_DATABASE_INFO ... 3106
PG_DEFAULT_ACL ... 3107
PG_EXTERNAL_SCHEMA ... 3110
PG_LIBRARY ... 3111
PG_PROC_INFO ... 3112
PG_STATISTIC_INDICATOR ... 3113
PG_TABLE_DEF ... 3114
PG_USER_INFO ... 3116
Consulta de las tablas de catálogos ... 3117
Referencia de la configuración ... 3124
Modificación de la configuración del servidor ... 3125
analyze_threshold_percent ... 3126
Valores (predeterminados en negrita) ... 3126
Descripción ... 3126
Ejemplos ... 3127
cast_super_null_on_error ... 3127
Valores (predeterminados en negrita) ... 3127
Descripción ... 3127
datashare_break_glass_session_var ... 3127
Valores (predeterminados en negrita) ... 3127
Descripción ... 3127
Ejemplo ... 3128

datestyle ..... 3128
Valores (predeterminados en negrita) ..... 3128
Descripción ..... 3127
Ejemplo ..... 3128
default_geometry_encoding ..... 3129
Valores (predeterminados en negrita) ..... 3129
Descripción ..... 3127
describe_field_name_in_uppercase ..... 3129
Valores (predeterminados en negrita) ..... 3129
Descripción ..... 3127
Ejemplo ..... 3128
downcase_delimited_identifier ..... 3130
Valores (predeterminados en negrita) ..... 3130
Descripción ..... 3127
Notas de uso ..... 3130
enable_case_sensitive_identifier ..... 3131
Valores (predeterminados en negrita) ..... 3131
Descripción ..... 3131
Ejemplos ..... 3132
Notas de uso ..... 3133
enable_case_sensitive_super_attribute ..... 3135
Valores (predeterminados en negrita) ..... 3135
Descripción ..... 3135
Ejemplos ..... 3135
Notas de uso ..... 3136
enable_numeric_rounding ..... 3137
Valores (predeterminados en negrita) ..... 3137
Descripción ..... 3137
Ejemplo ..... 3137
enable_result_cache_for_session ..... 3139
Valores (predeterminados en negrita) ..... 3139
Descripción ..... 3139
Ejemplo ..... 3139
enable_vacuum_boost ..... 3139
Valores (predeterminados en negrita) ..... 3139
Descripción ..... 3127

error_on_nondeterministic_update ..... 3140
Valores (predeterminados en negrita) ..... 3140
Descripción ..... 3127
Ejemplo ..... 3128
extra_float_digits ..... 3140
Valores (predeterminados en negrita) ..... 3140
Descripción ..... 3141
Ejemplo ..... 3141
interval_forbid_composite_literals ..... 3142
Valores (predeterminados en negrita) ..... 3142
Descripción ..... 3127
json_serialization_enable ..... 3143
Valores (predeterminados en negrita) ..... 3143
Descripción ..... 3127
json_serialization_parse_nested_strings ..... 3143
Valores (predeterminados en negrita) ..... 3143
Descripción ..... 3127
max_concurrency_scaling_clusters ..... 3144
Valores (predeterminados en negrita) ..... 3144
Descripción ..... 3144
max_cursor_result_set_size ..... 3144
Valores (predeterminados en negrita) ..... 3144
Descripción ..... 3144
mv_enable_aqmv_for_session ..... 3144
Valores (predeterminados en negrita) ..... 3144
Descripción ..... 3144
navegate_super_null_on_error ..... 3145
Valores (predeterminados en negrita) ..... 3145
Descripción ..... 3127
parse_super_null_on_error ..... 3145
Valores (predeterminados en negrita) ..... 3145
Descripción ..... 3127
pg_federation_repeatable_read ..... 3145
Valores (predeterminados en negrita) ..... 3145
Descripción ..... 3127
Ejemplos ..... 3146

query_group ..... 3146
Valores (predeterminados en negrita) ..... 3146
Descripción ..... 3146
search_path ..... 3147
Valores (predeterminados en negrita) ..... 3147
Descripción ..... 3147
Ejemplo ..... 3148
spectrum_enable_pseudo_columns ..... 3149
Valores (predeterminados en negrita) ..... 3149
Descripción ..... 3149
Ejemplo ..... 3149
enable_spectrum_oid ..... 3150
Valores (predeterminados en negrita) ..... 3150
Descripción ..... 3150
Ejemplo ..... 3150
spectrum_query_maxerror ..... 3150
Valores (predeterminados en negrita) ..... 3150
Descripción ..... 3150
Ejemplo ..... 3150
statement_timeout ..... 3151
Valores (predeterminados en negrita) ..... 3151
Descripción ..... 3151
Ejemplo ..... 3151
stored_proc_log_min_messages ..... 3152
Valores (predeterminados en negrita) ..... 3152
Descripción ..... 3127
timezone ..... 3152
Valores (predeterminados en negrita) ..... 3152
Sintaxis ..... 3152
Descripción ..... 3153
Formatos de zonas horarias ..... 3153
Ejemplos ..... 3155
use_fips_ssl ..... 3156
Valores (predeterminados en negrita) ..... 3156
Descripción ..... 3127
wlm_query_slot_count ..... 3157

Valores (predeterminados en negrita) ..... 3157
Descripción ..... 3157
Ejemplos ..... 3158
Historial de documentos ..... 3159
Actualizaciones anteriores ..... 3170

# Introducción a Amazon Redshift

Le damos la bienvenida a la Guía para desarrolladores de bases de datos de Amazon Redshift. Esta guía se centra en ayudarlo a comprender cómo usar Amazon Redshift para crear y administrar un almacenamiento de datos. Si trabaja con bases de datos como diseñador, desarrollador de software o administrador, le proporciona la información que necesita para diseñar, desarrollar, hacer consultas y mantener el almacenamiento de datos.

Amazon Redshift es un servicio de almacenamiento de datos administrado a escala de petabytes en la nube . Amazon Redshift sin servidor le permite acceder a los datos y analizarlos sin todas las configuraciones habituales de un almacenamiento de datos aprovisionado. Los recursos se aprovisionan automáticamente y la capacidad del almacenamiento de datos se escala de forma inteligente para ofrecer un rendimiento rápido incluso para las cargas de trabajo más exigentes e impredecibles. No incurrirá en gastos cuando el almacenamiento de datos esté inactivo, por lo que solo pagará por lo que utilice. Independientemente del tamaño del conjunto de datos, puede cargar datos y comenzar a realizar consultas de inmediato en el editor de consultas de Amazon Redshift v2 o en su herramienta de inteligencia empresarial (BI) favorita. Disfrute de la mejor relación precio-rendimiento y de las conocidas características de SQL en un entorno sin administración y fácil de utilizar.

Temas


- Requisitos previos para utilizar Amazon Redshift
- Arquitectura de Amazon Redshift
- Base de datos de muestra


## Requisitos previos para utilizar Amazon Redshift

En este tema se describen los requisitos previos que necesita para utilizar Amazon Redshift.

Antes de utilizar esta guía, debe leer Introducción a los almacenamientos de datos de Amazon Redshift sin servidor, donde se explica cómo realizar las siguientes tareas.

- Crear un almacenamiento de datos con Amazon Redshift sin servidor.
- Carga de datos de muestra con el editor de consultas de Amazon Redshift v2
- Carga de datos desde Amazon S3.

También debe saber cómo utilizar el cliente SQL y debe contar con los conocimientos fundamentales del lenguaje SQL.

# Arquitectura de Amazon Redshift

Este tema lo ayuda a comprender los componentes que integran Amazon Redshift.

Un almacenamiento de datos de Amazon Redshift es un sistema de administración y consulta de bases de datos relacionales de clase empresarial.

Amazon Redshift admite las conexiones de clientes con muchos tipos de aplicaciones, incluidas las herramientas de análisis, datos, generación de informes e inteligencia empresarial (BI).

Cuando ejecuta consultas de análisis, está recuperando, comparando y evaluando grandes cantidades de datos en operaciones de varias etapas para producir un resultado final.

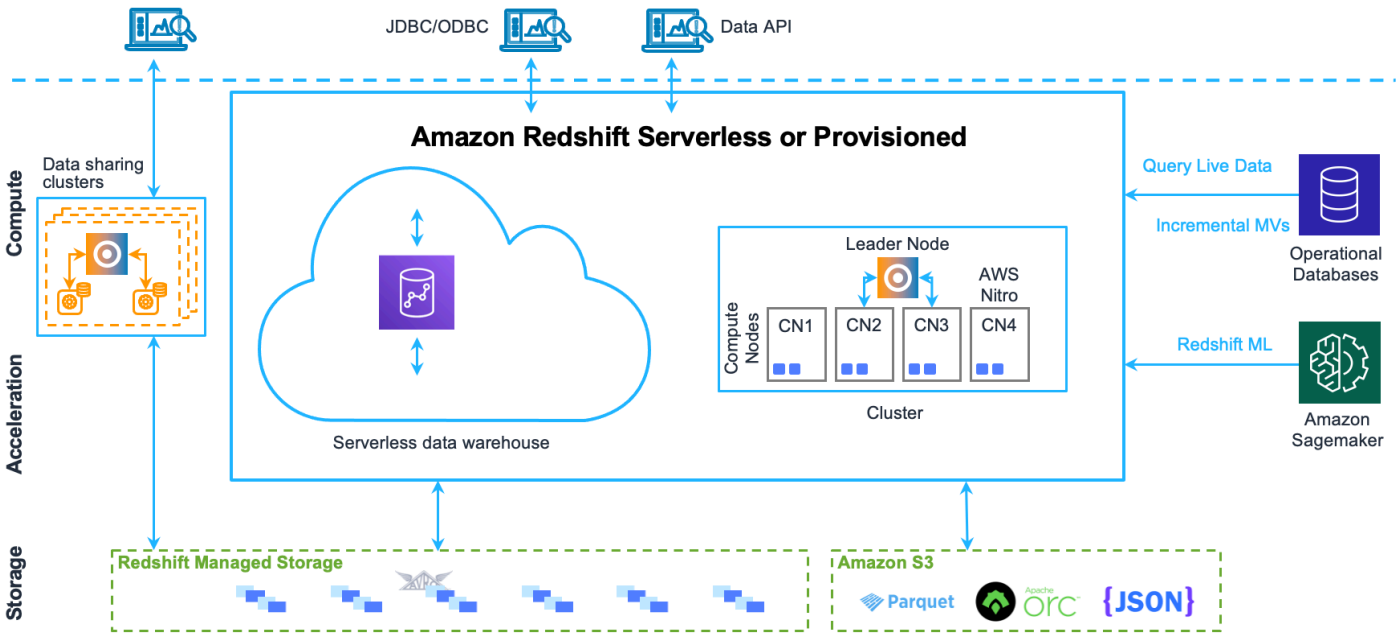
Amazon Redshift logra un almacenamiento eficiente y un rendimiento óptimo de las consultas mediante la combinación del procesamiento masivo en paralelo, el almacenamiento de datos en columnas y los esquemas de codificación para la compresión muy eficiente y orientada de datos. En esta sección, se presenta una introducción a la arquitectura del sistema de Amazon Redshift.

Temas


- Arquitectura del sistema de almacenamiento de datos
- Rendimiento de Amazon Redshift
- Almacenamiento en columnas
- Administración de la carga de trabajo
- Uso de Amazon Redshift con otros servicios


## Arquitectura del sistema de almacenamiento de datos

En esta sección, se explican los componentes de los que consta la arquitectura de almacenamiento de datos de Amazon Redshift, como se puede ver en la siguiente imagen.




## Aplicaciones cliente

Amazon Redshift se integra a diversas herramientas de carga de datos y ETL (extracción, transformación y carga) y a diversas herramientas de inteligencia empresarial (BI), generación de informes, minería de datos y análisis. Amazon Redshift se basa en el estándar abierto PostgreSQL, por lo que la mayoría de las aplicaciones cliente SQL existentes funcionarán con solo un mínimo de cambios. Para obtener más información acerca de las importantes diferencias que existen entre Amazon Redshift SQL y PostgreSQL, consulte Amazon Redshift y PostgreSQL.

## Clústeres

El principal componente de la infraestructura de un almacenamiento de datos de Amazon Redshift es el clúster.

Un clúster se compone de uno o varios nodos de computación. Si un clúster se aprovisiona con dos o más nodos de computación, un nodo principal adicional coordina los nodos de computación y administra la comunicación externa. La aplicación cliente interactúa de forma directa solo con el nodo principal. Los nodos de computación son transparentes para las aplicaciones externas.

## Nodo principal

El nodo principal administra las comunicaciones con los programas de clientes y todas las comunicaciones con los nodos de computación. Analiza y desarrolla los planes de ejecución para realizar las operaciones de bases de datos, en particular, la serie de pasos necesarios para obtener

resultados de consultas complejas. Según el plan de ejecución, el nodo principal compila un código, lo distribuye a los nodos de computación y les asigna una parte de los datos a cada uno.

El nodo principal distribuye instrucciones SQL a los nodos de computación solo cuando una consulta hace referencia a tablas que se encuentran almacenadas en los nodos de computación. Todas las demás consultas se ejecutan exclusivamente en el nodo principal. Amazon Redshift está diseñado para implementar ciertas funciones SQL solo en el nodo principal. Una consulta que utiliza cualquiera de estas funciones devolverá un mensaje de error si hace referencia a tablas que residen en los nodos de computación. Para obtener más información, consulte Funciones SQL admitidas en el nodo principal.

### Nodos de computación

El nodo principal compila un código para los elementos individuales del plan de ejecución y lo asigna a los nodos de computación individuales. Los nodos de computación ejecutan el código compilado y envían resultados intermedios de vuelta al nodo principal para su agregación final.

Cada nodo de computación tiene su propia CPU y memoria dedicadas, que se determinan por el tipo de nodo. A medida que la carga de trabajo crece, puede aumentar la capacidad de computación de un clúster mediante el incremento del número de nodos, la actualización del tipo de nodo o ambas acciones.

Amazon Redshift proporciona varios tipos de nodo que se ajustan a sus necesidades de computación. Para obtener información detallada acerca de cada tipo de nodo, consulte Clústeres de Amazon Redshift en la Guía de administración de Amazon Redshift.

### Redshift Managed Storage

Los datos del almacenamiento de datos se almacenan en Redshift Managed Storage (RMS) de nivel de almacenamiento independiente. RMS ofrece la posibilidad de escalar el almacenamiento hasta petabytes mediante el almacenamiento de Amazon S3. RMS le permite escalar y pagar la computación y el almacenamiento de forma independiente, de modo que pueda dimensionar su clúster en función únicamente de sus necesidades de computación. Utiliza automáticamente el almacenamiento local basado en SSD de alto rendimiento como caché de nivel 1. También aprovecha las optimizaciones, como la temperatura del bloque de datos, la antigüedad del bloque de datos y los patrones de cargas de trabajo para ofrecer un alto rendimiento y, al mismo tiempo, escalar el almacenamiento automáticamente a Amazon S3 cuando sea necesario sin necesidad de realizar ninguna acción.

### Sectores del nodo

Un nodo de computación está particionado en sectores. A cada sector se le asigna una parte de la memoria y del espacio en disco del nodo, donde se procesa una parte de la carga de trabajo asignada al nodo. El nodo principal administra los datos de distribución a los sectores y les reparte la carga de trabajo de cualquier consulta u otra operación de base de datos. A continuación, los sectores funcionan en paralelo para completar la operación.

El número de sectores por nodo está determinado por el tamaño de nodo del clúster. Para obtener más información acerca del número de sectores para cada tamaño de nodo, consulte Acerca de clústeres y nodos en la Guía de administración de Amazon Redshift.

Cuando crea una tabla, opcionalmente puede especificar una columna como la clave de distribución. Cuando se carga la tabla con los datos, las filas se distribuyen a los sectores del nodo de acuerdo con la clave de distribución que se defina para una tabla. La elección de una buena clave de distribución permite que Amazon Redshift utilice el procesamiento en paralelo para cargar datos y ejecutar consultas de forma eficiente. Para obtener más información acerca de una clave de distribución, consulte Elegir el modo de distribución recomendado.

Red interna

Amazon Redshift aprovecha las conexiones de ancho de banda alto, la gran proximidad y los protocolos de comunicación personalizados para proporcionar una comunicación de red privada de muy alta velocidad entre el nodo principal y los nodos informáticos. Los nodos de computación se ejecutan en una red aislada independiente a las que las aplicaciones cliente nunca obtienen acceso directamente.

Bases de datos

Un clúster contiene una o varias bases de datos. Los datos de usuario se almacenan en los nodos de computación. El cliente SQL se comunica con el nodo principal y este coordina la ejecución de consultas con los nodos de computación.

Amazon Redshift es un sistema de administración de base de datos relacional (RDBMS), por lo que es compatible con otras aplicaciones de RDBMS. Aunque proporciona la misma funcionalidad que un RDBMS típico, incluidas las funciones de procesamiento de transacciones online (OLTP), como insertar y eliminar datos, Amazon Redshift está optimizado para realizar informes y análisis de alto rendimiento de conjuntos de datos muy grandes.

Amazon Redshift se basa en PostgreSQL. Amazon Redshift y PostgreSQL tienen una serie de diferencias muy importantes que debe tener en cuenta al diseñar y desarrollar aplicaciones de

almacenamiento de datos. Para obtener más información acerca de las diferencias entre Amazon Redshift SQL y PostgreSQL, consulte Amazon Redshift y PostgreSQL.

# Rendimiento de Amazon Redshift

En este tema se describen los componentes de Amazon Redshift que impulsan el rendimiento. Comprender estos componentes lo ayudará a ajustar el rendimiento y a solucionar problemas de rendimiento deficiente con Amazon Redshift.

Amazon Redshift logra una ejecución de consultas ultrarrápida gracias a las siguientes características de rendimiento.

Temas


- Procesamiento masivo en paralelo
- Almacenamiento de datos en columnas
- Compresión de datos
- Optimizador de consultas
- Almacenamiento en caché de los resultados
- Código compilado


## Procesamiento masivo en paralelo

El procesamiento masivo en paralelo (MPP) permite una ejecución rápida de las consultas más complejas que funcionan en grandes cantidades de datos. Varios nodos de computación administran todo el procesamiento de consultas que dirige a la agregación de resultado final con cada núcleo de cada nodo que ejecuta los mismos segmentos de consulta compilados en partes de todos los datos.

Amazon Redshift distribuye las filas de una tabla entre los nodos informáticos para que los datos se puedan procesar en paralelo. Al seleccionar una clave de distribución adecuada para cada tabla, puede optimizar la distribución de datos para equilibrar la carga de trabajo y minimizar la transferencia de datos de un nodo a otro. Para obtener más información, consulte Elegir el modo de distribución recomendado.

La carga de datos de archivos sin formato aprovecha el procesamiento en paralelo mediante la distribución de la carga de trabajo entre varios nodos mientras lee varios archivos simultáneamente. Para obtener más información acerca de la carga de datos en las tablas, consulte Prácticas recomendadas de Amazon Redshift para la carga de datos.

## Almacenamiento de datos en columnas

El almacenamiento en columnas para tablas de bases de datos reduce de forma drástica los requisitos generales de E/S de disco y es un factor importante en la optimización del rendimiento analítico de consultas. Para obtener más información, consulte Almacenamiento en columnas.

Cuando las columnas se ordenan adecuadamente, el procesador de consultas puede filtrar de forma rápida un subconjunto de bloques de datos grande. Para obtener más información, consulte Elección de la clave de clasificación recomendada.

## Compresión de datos

La compresión de datos reduce los requisitos de almacenamiento, lo que permite disminuir la E/S del disco y mejorar el rendimiento de las consultas. Cuando se ejecuta una consulta, los datos comprimidos se leen en la memoria y luego se descomprimen durante la ejecución de la consulta. Cargar menos datos en la memoria permite a Amazon Redshift asignar más memoria al análisis de datos. Como el almacenamiento en columnas guarda los datos similares de forma secuencial, Amazon Redshift puede aplicar codificaciones de compresión adaptables que estén asociadas específicamente a los tipos de datos en columnas. La mejor forma de habilitar la compresión de datos en las columnas de las tablas es permitiendo que Amazon Redshift aplique las codificaciones de compresión óptimas cuando se carga la tabla con los datos. Para obtener más información acerca del uso de la compresión de datos automática, consulte Carga de tablas con compresión automática.

## Optimizador de consultas

El motor de ejecución de consultas de Amazon Redshift incorpora un optimizador de consultas que tiene en cuenta el procesamiento masivo en paralelo (MPP) y también aprovecha el almacenamiento de datos orientado en columnas. El optimizador de consultas de Amazon Redshift implementa mejoras y extensiones importantes para procesar consultas de análisis complejas que suelen incluir uniones, consultas secundarias y agrupaciones de varias tablas. Para obtener más información acerca de la optimización de consultas, consulte Ajuste del rendimiento de las consultas.

## Almacenamiento en caché de los resultados

Para reducir el tiempo de ejecución de las consultas y mejorar el rendimiento del sistema, Amazon Redshift almacena en caché los resultados de ciertos tipos de consultas en la memoria del nodo principal. Cuando un usuario envía una consulta, Amazon Redshift busca en la caché de resultados una copia válida de los resultados de la consulta. Si se encuentra alguna coincidencia en la caché de

resultados, Amazon Redshift utiliza estos resultados almacenados en caché y no ejecuta la consulta. El almacenamiento en caché de los resultados es transparente para el usuario.

El almacenamiento en la caché de resultados está activado de forma predeterminada. Para deshabilitar el almacenamiento en la caché de resultados para la sesión actual, establezca el parámetro enable_result_cache_for_session en `off`.

Amazon Redshift utiliza los resultados almacenados en caché para una nueva consulta cuando se cumplen todas las condiciones siguientes:

- El usuario que envía la consulta tiene permisos de acceso para los objetos utilizados en ella.
- La tabla o las vistas de la consulta no se han modificado.
- La consulta no utiliza una función que debe evaluarse cada vez que se ejecuta, como GETDATE.
- La consulta no referencia las tablas externas de Amazon Redshift Spectrum.
- Los parámetros de configuración que podrían afectar a los resultados de la consulta no se han modificado.
- La consulta coincide sintácticamente con la consulta existente en la memoria caché.

Para maximizar la eficacia de la caché y la eficiencia en el uso de los recursos, Amazon Redshift no almacena en caché algunos conjuntos de resultados de consultas de gran tamaño. Amazon Redshift determina si va a almacenar en caché o no los resultados de las consultas en función de diferentes factores. Estos factores incluyen el número de entradas existentes en la memoria caché y el tipo de instancia del clúster de Amazon Redshift.

Para determinar si una consulta ha utilizado la caché de resultados, consulte la vista SVL_QLOG del sistema. Si una consulta ha utilizado la caché de resultados, la columna source_query devuelve el ID de la consulta de origen. Si no se ha utilizado la caché de resultados, el valor de la columna source_query es NULL.

En el siguiente ejemplo, se muestra que las consultas enviadas por userid 104 y userid 102 utilizan la caché de resultados de las consultas ejecutadas por userid 100.

```
select userid, query, elapsed, source_query from svl_qlog
where userid > 1
order by query desc;

userid | query  | elapsed  | source_query
```

```
-------+--------+----------+-------------
   104 | 629035 |       27 |       628919
   104 | 629034 |       60 |       628900
   104 | 629033 |       23 |       628891
   102 | 629017 |  1229393 |
   102 | 628942 |       28 |       628919
   102 | 628941 |       57 |       628900
   102 | 628940 |       26 |       628891
   100 | 628919 | 84295686 |
   100 | 628900 | 87015637 |
   100 | 628891 | 58808694 |
```

### Código compilado

El nodo principal distribuye el código compilado totalmente optimizado entre todos los nodos de un clúster. La compilación de consultas reduce los costos asociados con intérpretes y, por lo tanto, aumenta la velocidad de ejecución, en especial para consultas complejas. El código compilado se almacena en caché y se comparte entre las sesiones del mismo clúster. Como resultado, las futuras ejecuciones de la misma consulta serán más rápidas, a menudo incluso con parámetros diferentes.

El motor de ejecución de consultas compila código diferente para los protocolos de conexión JDBC y ODBC, de este modo, dos clientes que utilicen protocolos diferentes incurren por separado en el costo inicial de compilación del código. Los clientes que usan el mismo protocolo, sin embargo, se beneficiarán del uso compartido del código en caché.

## Almacenamiento en columnas

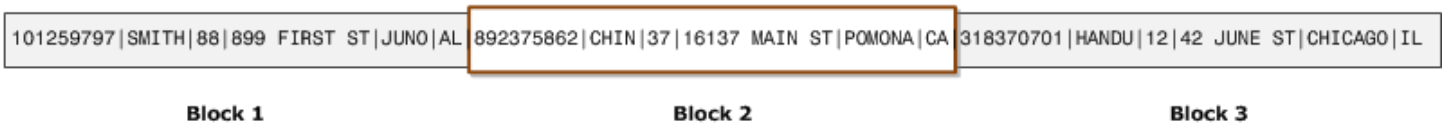
En esta sección se describe el almacenamiento en columnas, que es el método que utiliza Amazon Redshift para almacenar datos tabulares de forma eficiente.

El almacenamiento en columnas para tablas de bases de datos es un factor importante en la optimización del rendimiento analítico de consultas, porque reduce drásticamente los requisitos generales de E/S de disco. Esto reduce la cantidad de datos que necesita cargar del disco.

En la siguiente serie de ejemplos se describe cómo el almacenamiento de datos en columnas incorpora eficiencia y cómo eso se traduce en mejoras cuando se recuperan datos en la memoria.

En el primer ejemplo se muestra cómo los registros de las tablas de bases de datos suelen almacenarse por filas en los bloques del disco.

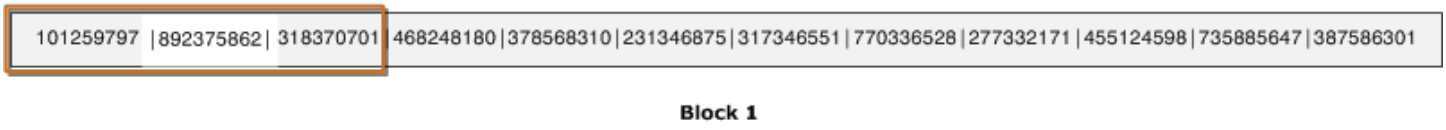
| SSN | Name | Age | Addr | City | St |
|---|---|---|---|---|---|
| 101259797 | SMITH | 88 | 899 FIRST ST | JUNO | AL |
| 892375862 | CHIN | 37 | 16137 MAIN ST | POMONA | CA |
| 318370701 | HANDU | 12 | 42 JUNE ST | CHICAGO | IL |




En una tabla de bases de datos relacionales típica, cada fila contiene valores de campo para un solo registro. En el almacenamiento de bases de datos en filas, los bloques de datos almacenan valores de forma secuencial en cada columna consecutiva que compone la fila completa. Si el tamaño del bloque es menor al tamaño de un registro, el almacenamiento de un registro completo puede ocupar más de un bloque. Si el tamaño del bloque es mayor al tamaño de un registro, el almacenamiento de un registro completo puede ocupar menos de un bloque, lo que se traduce en un uso ineficaz del espacio en disco. En las aplicaciones de procesamiento de transacciones online (OLTP), la mayoría de las transacciones con frecuencia incluyen la lectura y escritura de todos los valores de los registros completos, por lo general de a un registro o un número pequeño de registros a la vez. Como resultado, el almacenamiento en filas es óptimo para las bases de datos OLTP.

En el siguiente ejemplo se muestra cómo el almacenamiento en columnas guarda de forma secuencial los valores para cada columna en los bloques del disco.

| SSN | Name | Age | Addr | City | St |
|---|---|---|---|---|---|
| 101259797 | SMITH | 88 | 899 FIRST ST | JUNO | AL |
| 892375862 | CHIN | 37 | 16137 MAIN ST | POMONA | CA |
| 318370701 | HANDU | 12 | 42 JUNE ST | CHICAGO | IL |




Al usar el almacenamiento en columnas, cada bloque de datos almacena los valores de una sola columna para varias filas. A medida que los registros ingresan al sistema, Amazon Redshift convierte de forma transparente los datos en almacenamiento en columnas para cada una de las columnas.

En este ejemplo simplificado, mediante el almacenamiento en columnas, cada bloque de datos contiene valores de campo para hasta el triple de registros que los que contiene el almacenamiento

basado en filas. Esto significa que la lectura del mismo número de valores de campo de las columnas para el mismo número de registros requiere un tercio de las operaciones de E/S en comparación con el almacenamiento en filas. En la práctica, la eficiencia de almacenamiento es aún mayor mediante la utilización de tablas con números de columnas muy grandes y una gran cantidad de filas.

Una ventaja agregada es que, dado que cada bloque contiene el mismo tipo de datos, los datos del bloque pueden utilizar un esquema de compresión seleccionado específicamente para el tipo de datos de las columnas, lo que reduce aún más la actividad de E/S y el espacio en el disco. Para obtener más información acerca de las codificaciones de compresión según los tipos de datos, consulte Codificaciones de compresión.

El ahorro de espacio para almacenar datos en el disco también lleva a la recuperación y posterior almacenamiento de esos datos en la memoria. Ya que muchas operaciones de bases de datos solo necesitan obtener acceso u operar a la vez en una columna o en un pequeño número de ellas, se puede ahorrar espacio en la memoria con solo recuperar los bloques para columnas que realmente se necesitan para una consulta. Cuando las transacciones OLTP normalmente incluyen la mayoría de las columnas o todas en una fila para un pequeño número de registros, las consultas de data warehouse por lo general leen solo algunas columnas para un número de filas muy grande. Esto significa que la lectura del mismo número de valores de campo de las columnas para el mismo número de filas requiere una fracción de las operaciones de E/S. Utiliza una fracción de la memoria que se necesitaría para procesar bloques por filas. En la práctica, la mejora en la eficacia es proporcionalmente mayor gracias a la utilización de tablas con números de columnas muy grandes y una gran cantidad de filas. Por ejemplo, suponga que una tabla contiene 100 columnas. Una consulta que utiliza cinco columnas solo necesitará leer alrededor del cinco por ciento de los datos contenidos en la tabla. El ahorro se repite posiblemente para miles de millones o incluso billones de registros de grandes bases de datos. En contraste, la base de datos en filas también leería los bloques que contienen las 95 columnas innecesarias.

Los tamaños típicos de los bloques de base de datos oscilan entre 2 KB y 32 KB. Amazon Redshift utiliza un tamaño de bloque de 1 MB, que es más eficiente y reduce aún más la cantidad de solicitudes de E/S necesaria para realizar cualquier carga de base de datos u otras operaciones que forman parte de la ejecución de consultas.

## Administración de la carga de trabajo

En esta sección se describe la administración de cargas de trabajo (WLM), que lo ayuda a entender cómo Amazon Redshift prepara y ejecuta las consultas.

La administración de la carga de trabajo (WLM) de Amazon Redshift permite las prioridades flexibles dentro de las cargas de trabajo, por lo que las consultas cortas y de ejecución rápida no quedarán bloqueadas en las colas detrás de las consultas de ejecución prolongada. Amazon Redshift crea colas de consultas en tiempo de ejecución en función de las clases de servicios, que definen los parámetros de configuración de diferentes tipos de colas, incluidas las colas del sistema interno y las colas accesibles por el usuario. Desde una perspectiva de usuario, una clase de servicios accesibles para el usuario y una cola son funcionalmente equivalentes. Por motivos de coherencia, esta documentación usa el término cola para referirse tanto a la clase de servicios accesibles por el usuario como a la cola en tiempo de ejecución.

Redshift ofrece una administración automática de las cargas de trabajo, denominada WLM automática, que está ajustada para manejar cargas de trabajo variables y es la recomendada de forma predeterminada. Con la WLM automática, Redshift determina la utilización de los recursos a medida que llegan las consultas y determina dinámicamente si debe ejecutarlas en el clúster principal, en un clúster de escalado de simultaneidad o enviar cada una a una cola. (Cuando las consultas están en cola, la WLM automática prioriza las consultas de menor duración). La WLM automática maximiza el rendimiento total para que usted pueda mantener recursos de almacenamiento de datos eficientes. Usted ejecuta cargas de trabajo sin tener que preocuparse por su tamaño ni por la forma de programarlas. La WLM automática es la predeterminada para los clústeres aprovisionados. Para obtener más información, consulte Implementación de WLM automática.

> **Note**
>
> Los grupos de trabajo de Amazon Redshift sin servidor utilizan siempre WLM automática.

Cuando se ejecutan muchas consultas o consultas que consumen muchos recursos, la administración de cargas de trabajo puede escalar a recursos de computación adicionales cuando las cargas de trabajo se ponen en cola en los recursos locales. El escalado de simultaneidad con WLM automática admite un rendimiento coherente para un número casi ilimitado de usuarios y consultas simultáneas.

Los clústeres aprovisionados de Redshift ofrecen WLM manual si necesita una optimización manual detallada. Aquí, el cliente administra la asignación de recursos, la simultaneidad de las consultas y las colas. Cuando se ejecuta una consulta, WLM la asigna a una cola según el grupo de usuarios del usuario o por coincidencia con un grupo de consulta que figure en la configuración de la cola.

Esto se configura con una etiqueta de grupo de consulta que establece el usuario. Para obtener más información, consulte Implementación de WLM manual.

Aunque la WLM manual puede ajustarse con el tiempo para coincidir con sus patrones de carga de trabajo, en la mayoría de los casos desaconsejamos utilizarla porque su naturaleza estática puede dificultarle la adaptación a las cargas de trabajo cambiantes a lo largo de un día o de un periodo prolongado. Requiere un mayor monitoreo y un ajuste continuo. Además, en muchos casos la WLM manual no utiliza los recursos de computación con la misma eficacia que la WLM automática, por ejemplo, si las colas se establecen manualmente para limitar la memoria que se les asigna.

Una métrica importante para medir el éxito de la configuración de la administración de la carga de trabajo es el rendimiento del sistema, es decir, cuántas consultas se completan correctamente. El rendimiento del sistema se mide en consultas por segundo. Para obtener más información sobre las métricas del sistema, consulte Monitoreo del rendimiento del clúster de Amazon Redshift.

La forma más sencilla de administrar la configuración de WLM es mediante la consola de administración de Amazon Redshift. También puede utilizar la interfaz de la línea de comandos (CLI) de Amazon Redshift o la API de Amazon Redshift. Para obtener más información acerca de cómo implementar y utilizar la administración de la carga de trabajo, consulte Implementación de la administración de la carga de trabajo.

## Uso de Amazon Redshift con otros servicios

En esta sección se describe cómo puede utilizar otros servicios como orígenes y destinos de los datos de Amazon Redshift.

Amazon Redshift se integra a otros servicios de AWS para permitirle transferir, transformar y cargar los datos de manera rápida y confiable usando características de seguridad de datos.

S3

Amazon Simple Storage Service (Amazon S3) es un servicio web que almacena datos en la nube. Amazon Redshift utiliza el procesamiento en paralelo para leer y cargar datos de varios archivos de datos almacenados en buckets de Amazon S3. Para obtener más información, consulte Carga de datos desde Amazon S3.

También puede usar el procesamiento en paralelo para exportar datos desde el almacenamiento de datos de Amazon Redshift hacia varios archivos de datos de Amazon S3. Para obtener más información, consulte Descarga de datos en Amazon Redshift.

DynamoDB

Amazon DynamoDB es un servicio de base de datos NoSQL completamente administrado. Puede usar el comando COPY para cargar una tabla de Amazon Redshift con datos de una tabla específica de Amazon DynamoDB. Para obtener más información, consulte Carga de datos desde una tabla de Amazon DynamoDB.

SSH

Puede usar el comando COPY en Amazon Redshift para cargar datos de uno o más alojamientos remotos, como clústeres de Amazon EMR, instancias de Amazon EC2 u otros equipos. COPY se conecta a los hosts remotos usando SSH y ejecuta los comandos en los alojamientos remotos para generar datos. Amazon Redshift admite varias conexiones simultáneas. El comando COPY lee y carga la salida de varias fuentes de hosts en paralelo. Para obtener más información, consulte Carga de datos desde hosts remotos.

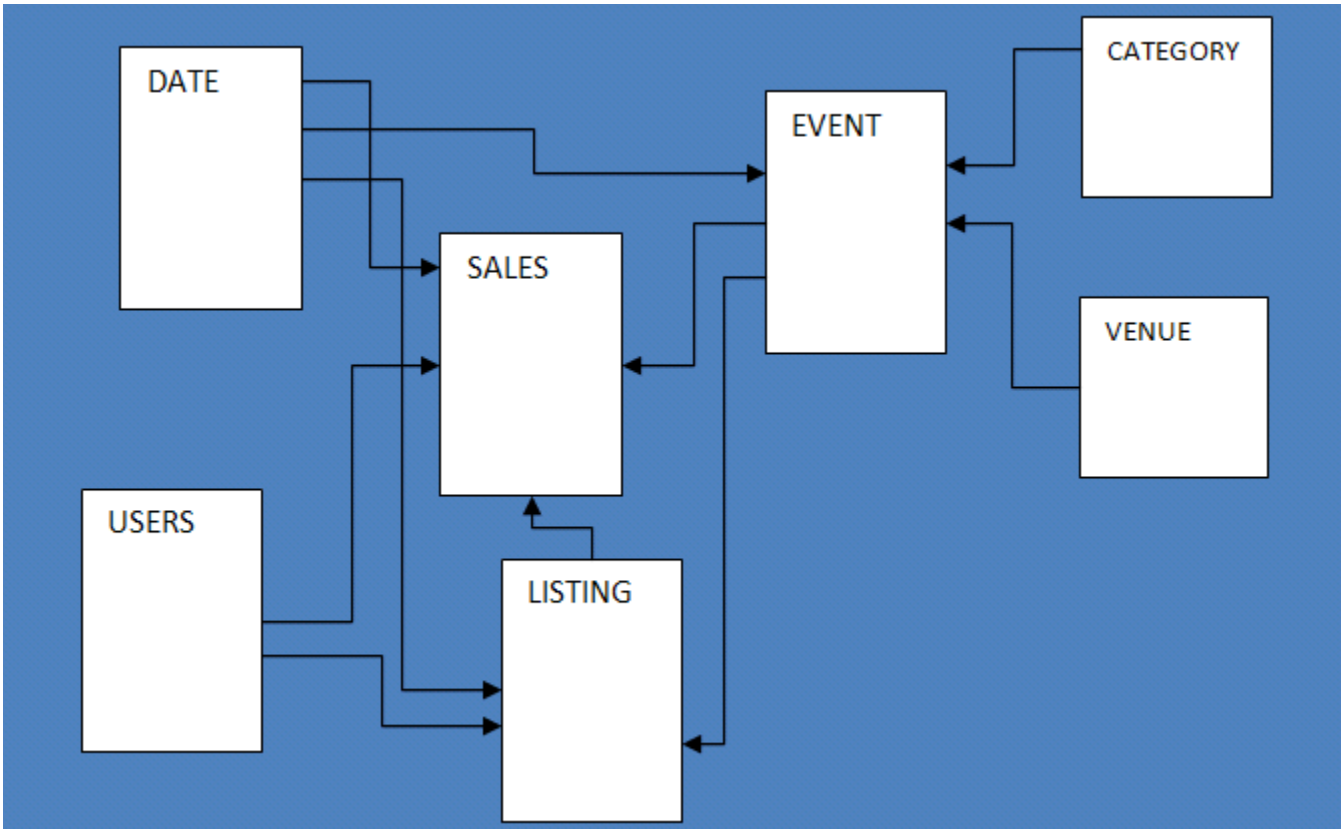
AWS DMS

Puede migrar datos a Amazon Redshift mediante AWS Database Migration Service. AWS DMS puede realizar la migración de los datos desde y hacia las bases de datos de código abierto y comerciales de mayor uso, como Oracle, PostgreSQL, Microsoft SQL Server, Amazon Redshift, clúster de Aurora DB, DynamoDB, Amazon S3, MariaDB y MySQL. Para obtener más información, consulte Uso de una base de datos de Amazon Redshift como destino de AWS Database Migration Service.

# Base de datos de muestra

En esta sección se describe TICKIT, una base de datos de muestra que se utiliza en los ejemplos de la documentación de Amazon Redshift.

Esa base de datos pequeña consta de siete tablas: dos de hechos y cinco de dimensiones. Puede cargar el conjunto de datos de TICKIT si sigue los pasos que se indican en Paso 4: Cargar datos desde Amazon S3 a Amazon Redshift en la Guía de introducción a Amazon Redshift.




Esta aplicación de base de datos de muestra ayuda a los analistas a realizar un seguimiento de la actividad de ventas del sitio web ficticio TICKIT, en el que los usuarios compran y venden en línea tickets para eventos deportivos, espectáculos y conciertos. En particular, los analistas pueden identificar los movimientos de tickets en el tiempo, las tasas de éxito de los vendedores y los eventos, lugares y temporadas más vendidos. Los analistas pueden usar esta información para ofrecer incentivos a los compradores y vendedores que frecuentan el sitio, a fin de atraer a nuevos usuarios e impulsar la publicidad y las promociones.

Por ejemplo, la siguiente consulta encuentra los cinco vendedores principales de San Diego, en función de la cantidad de tickets vendidos en 2008:

```
select sellerid, username, (firstname ||' '|| lastname) as name,
city, sum(qtysold)
from sales, date, users
where sales.sellerid = users.userid
and sales.dateid = date.dateid
and year = 2008
and city = 'San Diego'
group by sellerid, username, name, city
order by 5 desc
```

```
limit 5;

sellerid | username |       name        |   city    | sum
---------+----------+-------------------+-----------+-----
49977 | JJK84WTE | Julie Hanson      | San Diego |  22
19750 | AAS23BDR | Charity Zimmerman | San Diego |  21
29069 | SVL81MEQ | Axel Grant        | San Diego |  17
43632 | VAG08HKW | Griffin Dodson    | San Diego |  16
36712 | RXT40MKU | Hiram Turner      | San Diego |  14
(5 rows)
```

La base de datos usada para los ejemplos de esta guía presenta un conjunto de datos pequeño; cada una de las dos tablas de hechos tiene menos de 200 000 filas y las dimensiones varían, de 11 filas en la tabla CATEGORY hasta, aproximadamente, 50 000 en la tabla USERS.

En particular, los ejemplos de base de datos de esta guía demuestran las características clave del diseño de tablas de Amazon Redshift:

- Distribución de datos
- Orden de datos
- Compresión en columnas

Para obtener información sobre los esquemas de las tablas de la base de datos TICKIT, elija las siguientes pestañas:

CATEGORY table

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| CATID | SMALLINT | Clave primaria, un valor de ID único para cada fila. Cada fila representa un tipo de evento específico para el cual se compran y venden tickets. |
| CATGROUP | VARCHAR (10) | Nombre descriptivo de un grupo de eventos, como **`Shows`** y **`Sports`**. |
| CATNAME | VARCHAR (10) | Nombre descriptivo abreviado de un tipo de eventos en un grupo, como **`Opera`**y **`Musicals`**. |

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| CATDESC | VARCHAR(50) | Nombre descriptivo más largo del tipo de evento, como **`Musical theatre`**. |

DATE table

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| DATEID | SMALLINT | Clave primaria, un valor de ID único para cada fila. Cada fila representa un día del año calendario. |
| CALDATE | FECHA | Fecha de calendario, como **`2008-06-24`** . |
| DAY | CHAR (3) | Día de la semana (abreviado), como **`SA`**. |
| WEEK | SMALLINT | Número de semana, como **`26`**. |
| MONTH | CHAR (5) | Nombre del mes (abreviado), como **`JUN`**. |
| QTR | CHAR (5) | Número de trimestre (de **`1`** a **`4`**). |
| YEAR | SMALLINT | Los cuatro dígitos del año (**`2008`**). |
| HOLIDAY | BOOLEAN | La bandera indica si el día es un día festivo (EE. UU.). |

EVENT table

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| EVENTID | INTEGER | Clave primaria, un valor de ID único para cada fila. Cada fila representa un evento diferente que se lleva a cabo en un lugar específico y a una hora específic a. |

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| VENUEID | SMALLINT | Referencia de clave externa a la tabla VENUE. |
| CATID | SMALLINT | Referencia de clave externa a la tabla CATEGORY. |
| DATEID | SMALLINT | Referencia de clave externa a la tabla DATE. |
| EVENTNAME | VARCHAR (200) | Nombre del evento, como **`Hamlet`** o **`La Traviata`**. |
| STARTTIME | MARCA DE TIEMPO | Fecha y hora de inicio del evento sin abreviar, como **`2008-10-10 19:30:00`** . |

VENUE table

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| VENUEID | SMALLINT | Clave primaria, un valor de ID único para cada fila. Cada fila representa un lugar específico donde se lleva a cabo un evento. |
| VENUENAME | VARCHAR (100) | Nombre del recinto, como **`Cleveland Browns Stadium`**. |
| VENUECITY | VARCHAR (30) | Nombre de la ciudad, como **`Cleveland`** . |
| VENUESTATE | CHAR (2) | Abreviatura de dos letras para el estado o la provincia (Estados Unidos y Canadá), como **`OH`**. |
| VENUESEATS | INTEGER | Cantidad máxima de asientos disponibles en el recinto, si se conoce, como **`73200`**. Para fines de demostración, esta columna presenta valores nulos y ceros. |

USERS table

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| USERID | INTEGER | Clave primaria, un valor de ID único para cada fila. Cada fila representa a un usuario registrado (comprador, vendedor o ambos) quien ha listado o comprado tickets para al menos un evento. |
| USERNAME | CHAR (8) | Nombre de usuario alfanumérico de 8 caracteres, como **`PGL08LJI`**. |
| FIRSTNAME | VARCHAR (30) | Nombre de pila del usuario, como **`Victor`**. |
| LASTNAME | VARCHAR (30) | Apellido del usuario, como **`Hernandez`** . |
| CITY | VARCHAR (30) | Ciudad natal del usuario, como **`Naperville`** . |
| STATE | CHAR (2) | El estado natal del usuario, como **`GA`**. |
| CORREO ELECTRÓNICO | VARCHAR (100) | Dirección de correo electrónico del usuario; esta columna contiene valores latinos aleatorios, como **`turpis@accumsanlaoreet.org`** . |
| PHONE | CHAR (14) | Número de teléfono de 14 caracteres del usuario, como **`(818) 765-4255`**. |
| LIKESPORTS, ... | BOOLEAN | Serie de 10 columnas diferentes que identifican lo que le gusta o no le gusta al usuario, con valores **`true`** y **`false`**. |

LISTING table

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| LISTID | INTEGER | Clave primaria, un valor de ID único para cada fila. Cada fila representa la lista de un lote de tickets para un evento específico. |
| SELLERID | INTEGER | Referencia de clave externa a la tabla USERS, que identifica al usuario que vende los tickets. |
| EVENTID | INTEGER | Referencia de clave externa a la tabla EVENT. |
| DATEID | SMALLINT | Referencia de clave externa a la tabla DATE. |
| NUMTICKETS | SMALLINT | Cantidad de tickets disponibles para la venta, como **2** o **20**. |
| PRICEPERTICKET | DECIMAL (8,2) | Precio fijo de una entrada individual, como **27.00** o **206.00**. |
| TOTALPRICE | DECIMAL (8,2) | El precio total para este listado (NUMTICKETS*PRICEPERTICKET). |
| LISTTIME | MARCA DE TIEMPO | Fecha y hora sin abreviar en que se publicó el listado, como **2008-03-18 07:19:35** . |

SALES table

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| SALESID | INTEGER | Clave primaria, un valor de ID único para cada fila. Cada fila representa la venta de uno o más tickets para un evento específico, según se ofrecen en un listado específico. |

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| LISTID | INTEGER | Referencia de clave externa a la tabla LISTING. |
| SELLERID | INTEGER | Referencia de clave externa a la tabla USERS (el usuario que vendió los tickets). |
| BUYERID | INTEGER | Referencia de clave externa a la tabla USERS (el usuario que compró los tickets). |
| EVENTID | INTEGER | Referencia de clave externa a la tabla EVENT. |
| DATEID | SMALLINT | Referencia de clave externa a la tabla DATE. |
| QTYSOLD | SMALLINT | Cantidad de entradas vendidas, de **1** a **8**. (Se pueden vender un máximo de 8 tickets en una única transacción). |
| PRICEPAID | DECIMAL (8,2) | El precio total pagado por las entradas, como **75.00** o **488.00**. El precio individual de un ticket es PRICEPAID/QTYSOLD. |
| COMMISSION | DECIMAL (8,2) | 15 % de comisión que el negocio obtiene de la venta, como **11.25** o **73.20**. El vendedor recibe el 85% del valor de PRICEPAID. |
| SALETIME | MARCA DE TIEMPO | Fecha y hora sin abreviar en que se finalizó la venta, como **2008-05-24 06:21:47** . |

# Prácticas recomendadas de Amazon Redshift

A continuación, encontrará prácticas recomendadas para la planificación de una prueba de concepto, el diseño de tablas, la carga de datos en tablas y la escritura de consultas en Amazon Redshift, así como un análisis de cómo trabajar con Amazon Redshift Advisor.

Amazon Redshift no es igual a otros sistemas de base de datos SQL. Para aprovechar completamente los beneficios de la arquitectura de Amazon Redshift, debe diseñar, crear y cargar las tablas de manera específica, con el fin de utilizar el procesamiento masivo en paralelo, el almacenamiento de datos en columnas y la compresión de datos en columnas. Si los tiempos de carga de datos y de ejecución de consultas superan lo esperado o si son más largos de lo que desea, quizás esté pasando por alto información clave.

Si es un desarrollador de bases de datos SQL con experiencia, le recomendamos encarecidamente revisar este tema antes de comenzar a desarrollar su almacenamiento de datos de Amazon Redshift.

Si se está iniciando en el desarrollo de bases de datos SQL, este tema no es el mejor lugar donde comenzar. Le recomendamos que primero consulte Ejecución de comandos para definir y utilizar una base de datos en el almacenamiento de datos en la Guía de introducción de Amazon Redshift y que pruebe los ejemplos.

En este tema se brinda información general sobre los principios de desarrollo más importantes, junto con consejos y ejemplos específicos, así como prácticas recomendadas para implementar esos principios. No hay una práctica única que pueda implementarse para cada aplicación. Evalúe todas las opciones antes de finalizar el diseño de una base de datos. Para obtener más información, consulte Optimización de tablas automática, Carga de datos en Amazon Redshift, Ajuste del rendimiento de las consultas y los capítulos de referencia.

Temas

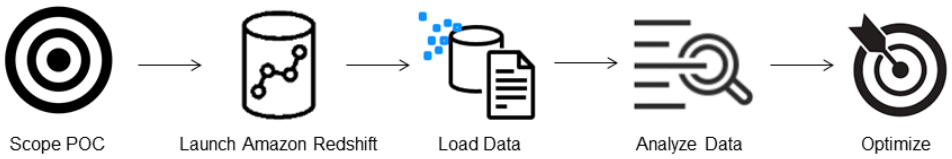

- Realización de una prueba de concepto (POC) de Amazon Redshift
- Prácticas recomendadas de Amazon Redshift para el diseño de tablas
- Prácticas recomendadas de Amazon Redshift para la carga de datos
- Prácticas recomendadas de Amazon Redshift para el diseño de consultas
- Seguimiento de las recomendaciones de Amazon Redshift Advisor

# Realización de una prueba de concepto (POC) de Amazon Redshift

Amazon Redshift es un popular almacenamiento de datos en la nube, que ofrece un servicio basado en la nube totalmente administrado que se integra con el lago de datos de Amazon Simple Storage Service de una organización, flujos en tiempo real, flujos de trabajo de machine learning (ML), flujos de trabajo transaccionales y mucho más. En las siguientes secciones, se le guiará a través del proceso de realización de una prueba de concepto (POC) en Amazon Redshift. Esta información le ayudará a establecer objetivos para su POC y a aprovechar las herramientas que pueden automatizar el aprovisionamiento y la configuración de los servicios para su POC.

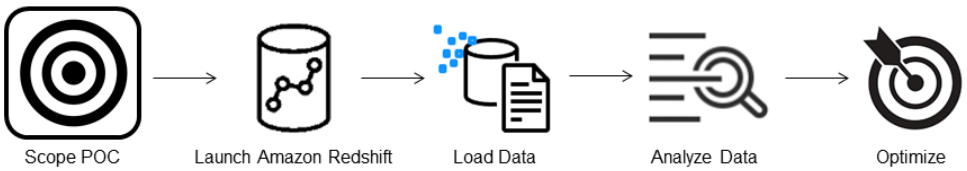
> **Note**
>
> Para obtener una copia de esta información en formato PDF, elija el enlace Ejecute su propio POC de Redshift en la página de recursos de Amazon Redshift.

Al realizar un POC de Amazon Redshift, usted prueba, comprueba y adopta características que van desde las mejores capacidades de seguridad de su clase, escalado elástico, fácil integración e ingestión hasta opciones flexibles de arquitecturas de datos descentralizadas.




Siga estos pasos para realizar el POC correctamente.

## Paso 1: defina el alcance del POC




A la hora de realizar un POC, puede optar por utilizar sus propios datos o puede usar conjuntos de datos de referencia. Cuando elige sus propios datos, ejecuta sus propias consultas en ellos. Con los

datos de evaluación de referencia, se proporcionan consultas de ejemplo con la referencia. Consulte Uso de conjuntos de datos de muestra para obtener más información si aún no está preparado para realizar una POC con sus propios datos.

En general, recomendamos utilizar dos semanas de datos para un POC de Amazon Redshift.

Para empezar, haga lo siguiente:

1. Identifique sus requisitos empresariales y funcionales y, a continuación, trabaje en sentido inverso. Los ejemplos más comunes son: un rendimiento más rápido, costos más bajos, probar una nueva carga de trabajo o característica, o comparar Amazon Redshift con otro almacenamiento de datos.
2. Establezca objetivos específicos que se conviertan en los criterios de éxito del POC. Por ejemplo, si busca un rendimiento más rápido, elabore una lista de los cinco procesos principales que desee acelerar e incluya los tiempos de ejecución actuales junto con el tiempo de ejecución requerido. Pueden ser informes, consultas, procesos de ETL, ingesta de datos o cualquiera que sea su punto débil actual.
3. Identifique el alcance específico y los artefactos necesarios para ejecutar las pruebas. ¿Qué conjuntos de datos necesita migrar o ingerir de forma continua en Amazon Redshift y qué consultas y procesos se necesitan para ejecutar las pruebas y compararlas con los criterios de éxito? Hay dos formas de hacer esto:

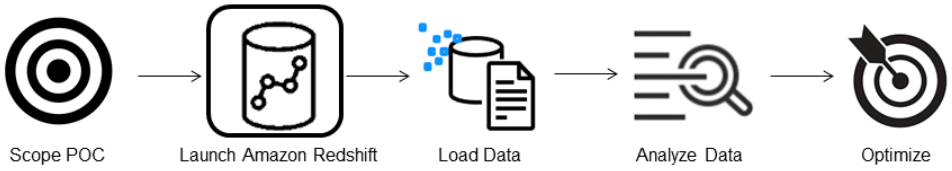
   Uso de sus propios datos

   - Para probar sus propios datos, elabore la lista mínima viable de artefactos de datos necesaria para probar sus criterios de éxito. Por ejemplo, si su almacenamiento de datos actual tiene 200 tablas, pero los informes que desea probar solo necesitan 20, su POC se puede ejecutar más rápido si utiliza solo el subconjunto de tablas más pequeño.

   Uso de conjuntos de datos de muestra

   - Si no tiene sus propios conjuntos de datos preparados, puede empezar a realizar un POC en Amazon Redshift utilizando los conjuntos de datos de referencia estándares del sector, como TPC-DS o TPC-H, y ejecutar ejemplos de consultas de referencia para aprovechar el potencial de Amazon Redshift. Se puede acceder a estos conjuntos de datos desde su almacenamiento de datos de Amazon Redshift después de crearlos. Para obtener instrucciones detalladas sobre cómo acceder a estos conjuntos de datos y ejemplos de consultas, consulte Paso 2: inicie Amazon Redshift.

# Paso 2: inicie Amazon Redshift




Amazon Redshift acelera la obtención de información con un almacenamiento de datos en la nube rápido, sencillo y seguro a escala. Puede empezar rápidamente iniciando su almacén en la consola de Redshift sin servidor y pasar de los datos a la información en cuestión de segundos. Con Redshift sin servidor, puede centrarse en lograr los resultados de su negocio sin preocuparse por la administración de su almacenamiento de datos.

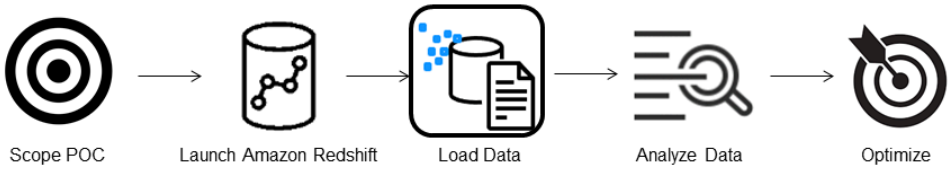
## Configuración de Amazon Redshift sin servidor

La primera vez que utilice Redshift sin servidor, la consola le guiará por los pasos necesarios para iniciar su almacén. También podría optar a un crédito para el uso de Redshift sin servidor en su cuenta. Para obtener más información sobre la elección de una prueba gratuita, consulte Prueba gratuita de Amazon Redshift. Siga los pasos de la sección Creación de un almacenamiento de datos con Amazon Redshift sin servidor de la Guía de introducción a Amazon Redshift para crear un almacenamiento de datos con Redshift sin servidor. Si no tiene un conjunto de datos que desee cargar, la guía también contiene los pasos para cargar un conjunto de datos de ejemplo.

Si ya ha iniciado Redshift sin servidor en su cuenta, siga los pasos que se indican en Creación de un grupo de trabajo con un espacio de nombres en la Guía de administración de Amazon Redshift. Cuando su almacén esté disponible, puede optar por cargar los datos de ejemplo disponibles en Amazon Redshift. Para obtener información sobre el uso del editor de consultas v2 de Amazon Redshift para cargar datos, consulte Carga de datos de ejemplo en la Guía de administración de Amazon Redshift.

Si va a traer sus propios datos en lugar de cargar el conjunto de datos de ejemplo, consulte Paso 3: cargue los datos.

# Paso 3: cargue los datos




Tras iniciar Redshift sin servidor, el siguiente paso es cargar los datos para el POC. Tanto si va a cargar un simple archivo CSV, ingiriendo datos semiestructurados desde S3 o transmitiendo en streaming datos directamente, Amazon Redshift proporciona la flexibilidad necesaria para trasladar los datos de forma rápida y sencilla a las tablas de Amazon Redshift desde el origen.

Elija uno de los siguientes métodos para cargar los datos.

## Carga de un archivo local

Para una ingesta y un análisis rápidos, puede utilizar el editor de consultas de Amazon Redshift v2 para cargar fácilmente los archivos de datos desde su escritorio local. Tiene la capacidad de procesar archivos en varios formatos, como CSV, JSON, AVRO, PARQUET, ORC y muchos más. Para que sus usuarios, como administradores, puedan cargar datos desde un escritorio local utilizando el editor de consultas v2, debe especificar un bucket común de Amazon S3 y la cuenta de usuario debe estar configurada con los permisos adecuados. Puede seguir el proceso de Data load made easy and secure in Amazon Redshift using Query Editor V2 para obtener instrucciones paso a paso.

## Carga de un archivo de Amazon S3

Para cargar datos desde un bucket de Amazon S3 en Amazon Redshift, comience por utilizar el comando COPY y especifique la ubicación de Amazon S3 de origen y la tabla de Amazon Redshift de destino. Asegúrese de que los permisos y los roles de IAM estén configurados correctamente para permitir el acceso de Amazon Redshift al bucket de Amazon S3 designado. Siga el Tutorial: carga de datos desde Amazon S3 para obtener instrucciones paso a paso. También puede elegir la opción Cargar datos en el editor de consultas v2 para cargar datos directamente desde su bucket de S3.

### Ingesta de datos continua

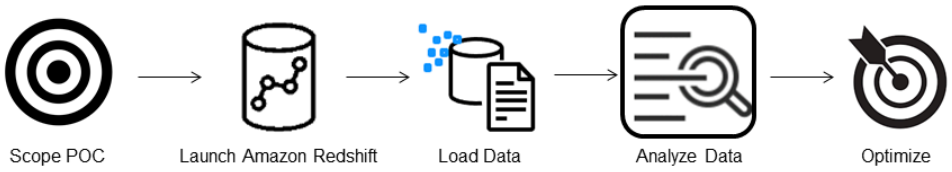
La copia automática (en versión preliminar) es una extensión del comando COPY y automatiza la carga continua de datos desde los buckets de Amazon S3. Al crear un trabajo de copia, Amazon Redshift detecta cuándo se crean nuevos archivos de Amazon S3 en una ruta especificada y, a continuación, los carga automáticamente sin su intervención. Amazon Redshift realiza un seguimiento de los archivos cargados para verificar que se cargan solo una vez. Para obtener instrucciones sobre cómo crear trabajos de copia, consulte COPY JOB

> **Note**
>
> La copia automática se encuentra actualmente en versión preliminar y solo se admite en clústeres aprovisionados en Regiones de AWS específicas. Para crear un clúster de la versión preliminar para la copia automática, consulte Creación de una integración de eventos de S3 para copiar automáticamente archivos desde buckets de Amazon S3.

### Carga de datos en streaming

La ingesta en streaming proporciona una ingesta de alta velocidad y baja latencia de datos de flujos de Amazon Kinesis Data Streams y Amazon Managed Streaming para Apache Kafka en Amazon Redshift. La ingesta en streaming de Amazon Redshift utiliza una vista materializada, que se actualiza directamente desde el flujo utilizando la actualización automática. La vista materializada se asigna al origen de datos del flujo. Puede realizar filtrados y agregaciones en los datos del flujo como parte de la definición de la vista materializada. Para obtener instrucciones paso a paso sobre cómo cargar datos desde un flujo, consulte Getting started with Amazon Kinesis Data Streams o Getting started with Amazon Managed Streaming for Apache Kafka.

## Paso 4: analice sus datos




Tras crear el grupo de trabajo y el espacio de nombres de Redshift sin servidor y cargar los datos, puede ejecutar consultas inmediatamente. Para ello, abra el editor de consultas v2 desde el panel

de navegación de la consola de Redshift sin servidor. Puede utilizar el editor de consultas v2 para probar la funcionalidad o el rendimiento de las consultas con sus propios conjuntos de datos.

## Consulta utilizando el editor de consultas de Amazon Redshift v2

Puede acceder al editor de consultas v2 desde la consola de Amazon Redshift. Consulte Simplify your data analysis with Amazon Redshift query editor v2 para obtener una guía completa sobre cómo configurar, conectar y ejecutar consultas con el editor de consultas v2.

Como alternativa, si desea ejecutar una prueba de carga como parte de su POC, puede hacerlo siguiendo estos pasos para instalar y ejecutar Apache JMeter.

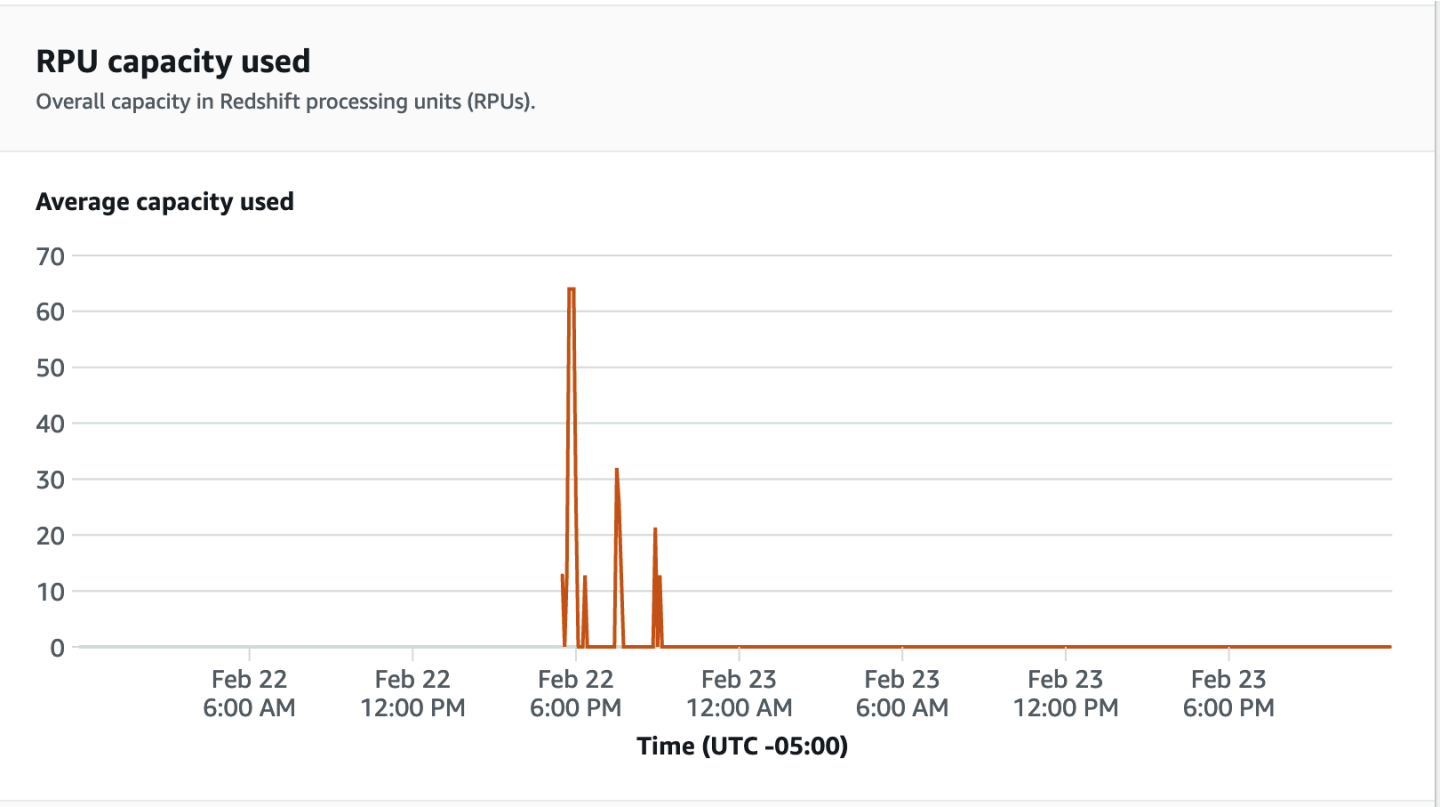
## Ejecución de una prueba de carga con Apache JMeter

Para realizar una prueba de carga que simule que “N” usuarios envían consultas de forma simultánea a Amazon Redshift, puede utilizar Apache JMeter, una herramienta de código abierto basada en Java.

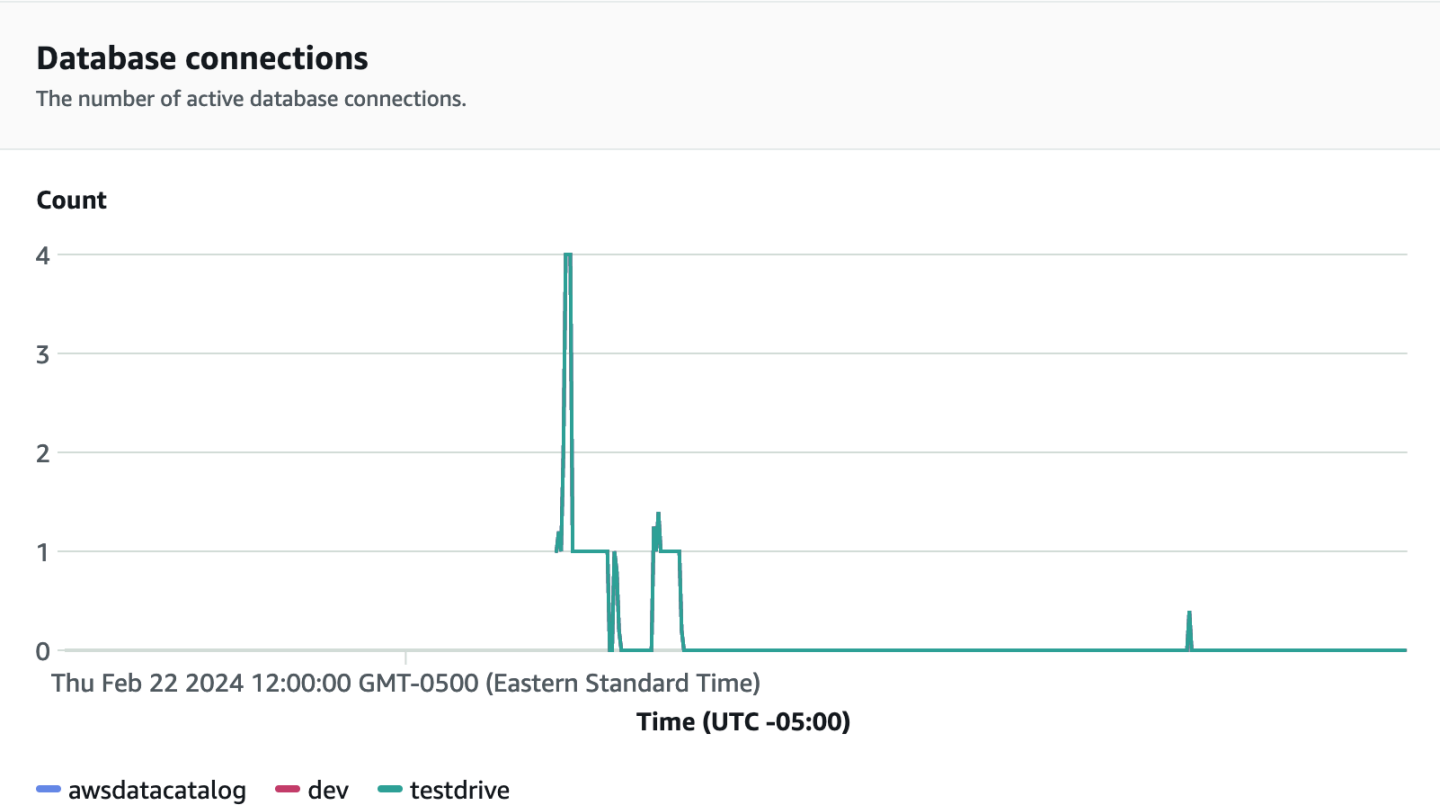
Para instalar y configurar Apache JMeter para que se ejecute en su grupo de trabajo de Redshift sin servidor, siga las instrucciones de Automate Amazon Redshift load testing with the AWS Analytics Automation Toolkit. Utiliza el kit de herramientas de automatización de análisis (AAA) de AWS, una utilidad de código abierto para implementar dinámicamente las soluciones de Redshift para iniciar automáticamente estos recursos. Si ha cargado sus propios datos en Amazon Redshift, asegúrese de realizar el Paso 5: personalice la opción SQL, para asegurarse de que suministra las instrucciones SQL adecuadas que desea probar en sus tablas. Pruebe cada una de estas instrucciones SQL una vez con el editor de consultas v2 para asegurarse de que se ejecutan sin errores.

Tras completar la personalización de las instrucciones SQL y finalizar el plan de pruebas, guarde y ejecute el plan de pruebas en su grupo de trabajo de Redshift sin servidor. Para monitorear el progreso de su prueba, abra la consola de Redshift sin servidor, navegue hasta Monitoreo de consultas y bases de datos, elija la pestaña Historial de consultas y visualice la información sobre sus consultas.

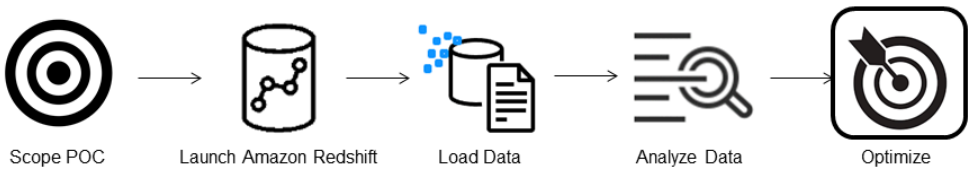
Para ver las métricas de rendimiento, elija la pestaña Rendimiento de la base de datos en la consola de Redshift sin servidor y podrá monitorear métricas como Conexiones de la base de datos y Uso de la CPU. Aquí puede ver un gráfico para monitorear la capacidad de la RPU utilizada y observar cómo Redshift sin servidor se escala automáticamente para satisfacer las demandas de carga de trabajo simultáneas mientras se ejecuta la prueba de carga en su grupo de trabajo.




Las conexiones de las bases de datos son otra métrica útil que se debe monitorear al ejecutar la prueba de carga para ver cómo su grupo de trabajo gestiona numerosas conexiones simultáneas en un momento dado para satisfacer las crecientes demandas de carga de trabajo.




## Paso 5: optimice




Amazon Redshift permite a decenas de miles de usuarios procesar exabytes de datos todos los días y potenciar sus cargas de trabajo de análisis al ofrecer una variedad de configuraciones y características para admitir casos de uso individuales. Al elegir entre estas opciones, los clientes buscan herramientas que les ayuden a determinar la configuración de almacenamiento de datos más óptimo para admitir su carga de trabajo de Amazon Redshift.

### Versiones de prueba

Puede utilizar Versiones de prueba para reproducir automáticamente su carga de trabajo actual en posibles configuraciones y analizar los resultados correspondientes para evaluar el destino óptimo al que migrar su carga de trabajo. Consulte Find the best Amazon Redshift configuration for your

workload using Redshift Test Drive para obtener información sobre el uso de Versiones de prueba para evaluar diferentes configuraciones de Amazon Redshift.

# Prácticas recomendadas de Amazon Redshift para el diseño de tablas

Cuando planifique la base de datos tenga en cuenta que algunas decisiones clave sobre el diseño de la tabla influyen considerablemente en el rendimiento general de la consulta. Las elecciones de diseño también tienen consecuencias importantes en los requisitos de almacenamiento, lo cual a su vez afecta el rendimiento de la consulta al reducir la cantidad de operaciones de E/S y reducen la memoria requerida para procesar las consultas.

En esta sección, encontrará un resumen de las decisiones de diseño más importantes y prácticas recomendadas para optimizar el rendimiento de las consultas. Optimización de tablas automática contiene explicaciones y ejemplos más detallados sobre las opciones de diseño de tablas.

Temas

- Elección de la clave de clasificación recomendada
- Elegir el modo de distribución recomendado
- Selección del comando COPY para las codificaciones de compresión
- Definición de restricciones de clave primaria y clave externa
- Uso del menor tamaño posible de columna
- Uso de tipos de datos de fecha/hora para columnas de fecha

## Elección de la clave de clasificación recomendada

Amazon Redshift almacena los datos en el disco en un determinado orden en función de la clave de ordenación. El optimizador de consultas de Amazon Redshift utiliza la ordenación cuando determina cuáles son los planes óptimos de consulta.

> **Note**
>
> Cuando utiliza la optimización automática de tablas, no necesita elegir la clave de ordenación de la tabla. Para obtener más información, consulte Optimización de tablas automática.

A continuación, se indican algunas sugerencias para lograr el mejor enfoque posible:

- Para que Amazon Redshift elija el orden adecuado, especifique AUTO en la clave de ordenación.
- Si los datos recientes se consultan con más frecuencia, especifique la columna de marca temporal como la columna inicial para la clave de ordenación.

  Las consultas son más eficientes debido a que pueden omitir bloques enteros que se encuentran fuera del intervalo de tiempo.
- Si utiliza filtros de igualdad o por rango con frecuencia en una columna, especifique esa columna como la clave de ordenación.

  Amazon Redshift puede omitir la lectura de bloques completos de datos de esa columna. Puede realizar esta operación porque realiza un seguimiento de los valores de columna mínimos y máximos almacenados en cada bloque y puede omitir los bloques que no se aplican al rango de predicados.
- Si combina tablas con frecuencia, especifique la columna de combinación como clave de ordenación y clave de distribución.

  Esto permite que el optimizador de consultas seleccione una combinación de fusión y ordenación, en lugar de una combinación hash más lenta. Debido a que los datos ya están ordenados en la clave de combinación, el optimizador de consultas puede omitir la fase de ordenación de la combinación de fusión y ordenación.

## Elegir el modo de distribución recomendado

Cuando ejecuta una consulta, el optimizador de consultas redistribuye las filas a los nodos de computación según se necesite para realizar combinaciones y agregaciones. El objetivo al seleccionar un estilo de distribución de tablas es reducir el impacto del paso de redistribución al localizar los datos en el lugar que deben estar antes de que se ejecute la consulta.

> **Note**
>
> Cuando utiliza la optimización automática de tablas, no necesita elegir el estilo de distribución de la tabla. Para obtener más información, consulte Optimización de tablas automática.

A continuación, se indican algunas sugerencias para lograr el mejor enfoque posible:

1. Distribuya la tabla de hechos y una tabla de dimensión en sus columnas comunes.

   Su tabla de hechos solo puede tener una clave de distribución. Ninguna de las tablas que se combinan con otra clave se ubica junto a la tabla de hechos. Seleccione una dimensión para colocar, según la frecuencia con que se combina y según el tamaño de las filas combinadas. Designe la clave principal de la tabla de dimensión y la clave externa correspondiente a la tabla de hechos como las claves DISTKEY.

2. Seleccione la mayor dimensión según el tamaño del conjunto de datos filtrado.

   Solo se deben distribuir las filas que se usan en la combinación; por lo tanto, considere el tamaño del conjunto de datos después del filtrado, no el tamaño de la tabla.

3. Seleccione una columna que tenga una cardinalidad alta en el conjunto de resultados filtrado.

   Por ejemplo, si distribuye una tabla de ventas en una columna de fecha, probablemente obtenga una distribución de fechas bastante uniforme, a menos que la mayoría de sus ventas se realicen por temporada. No obstante, si suele utilizar un predicado de rango restringido para filtrar un periodo de fechas limitado, la mayoría de las filas filtradas están en un conjunto limitado de secciones y la carga de trabajo de la consulta está descompensada.

4. Cambie algunas tablas de dimensión para que utilicen la distribución ALL.

   Si no se puede colocar una tabla de dimensión junto con la tabla de hechos u otra tabla de combinación de importancia, puede mejorar el rendimiento de las consultas de forma significativa distribuyendo la tabla completa a todos los nodos. El uso de la distribución ALL multiplica los requisitos de espacio de almacenamiento y aumenta los tiempos de carga y las operaciones de mantenimiento, por lo que debe analizar todos los factores antes de seleccionar la distribución ALL.

Para que Amazon Redshift elija el estilo de distribución adecuado, especifique `AUT0` en el estilo de distribución.

Para obtener más información acerca de la elección de estilos de distribución, consulte Distribución de datos para la optimización de consultas.

## Selección del comando COPY para las codificaciones de compresión

Cuando crea una tabla, puede especificar codificaciones de compresión, pero en la mayoría de los casos la compresión automática produce mejores resultados.

ENCODE AUTO es la opción predeterminada para las tablas. Cuando una tabla se establece a ENCODE AUTO, Amazon Redshift administra automáticamente la codificación de compresión para todas las columnas de la tabla. Para obtener más información, consulte CREATE TABLE y ALTER TABLE.

El comando COPY analiza los datos y aplica las codificaciones de compresión en una tabla vacía de forma automática como parte de la operación de carga.

La compresión automática equilibra el rendimiento general al elegir las codificaciones de compresión. Los exámenes de rango restringido podrían tener un rendimiento deficiente cuando las columnas de clave con ordenación se comprimen mucho más que otras columnas en la misma consulta. Como resultado la compresión automática elige una codificación de compresión menos eficiente para mantener equilibradas las columnas con clave de ordenación con las demás columnas.

Supongamos que la clave de ordenación de la tabla es una fecha o una marca temporal y que la tabla utiliza muchas columnas varchar grandes. En dicho caso es posible que el rendimiento sea mejor si no comprime la columna de clave de ordenación. Ejecute el comando ANALYZE COMPRESSION en la tabla y, luego, utilice las codificaciones para crear una tabla nueva, pero deje la codificación de compresión para la clave de ordenación.

Existe un costo de rendimiento para la codificación de compresión automática, pero solo si la tabla está vacía y si no tiene una codificación de compresión. En el caso de las tablas con una vida útil corta o las tablas que crea con frecuencia, como las tablas provisionales, cargue la tabla una vez con la compresión automática o ejecute el comando ANALYZE COMPRESSION. Utilice estas codificaciones para crear nuevas tablas. Puede agregar las codificaciones a la instrucción CREATE TABLE o utilizar CREATE TABLE LIKE para crear una tabla nueva con la misma codificación.

Para obtener más información, consulte Carga de tablas con compresión automática.

## Definición de restricciones de clave primaria y clave externa

Define las restricciones de clave principal y clave externa entre tablas siempre que corresponda. Aunque solo sean informativas, el optimizador de consultas utiliza esas restricciones para generar planes de consulta más eficientes.

No defina restricciones para la clave primaria ni la clave externa, a menos que la aplicación imponga dichas restricciones. Amazon Redshift no impone restricciones exclusivas, de clave primaria ni de clave externa.

Consulte Restricciones de tablas para obtener más información sobre la manera en que Amazon Redshift utiliza las restricciones.

### Uso del menor tamaño posible de columna

No se acostumbre a usar el mayor tamaño de columna por comodidad.

En lugar de eso, considere los valores más grandes que vaya a almacenar en las columnas y dimensiónelos adecuadamente. Por ejemplo, una columna CHAR para almacenar abreviaturas de estados o territorios de EE. UU. usada por la oficina de correos solo necesita ser CHAR(2).

### Uso de tipos de datos de fecha/hora para columnas de fecha

Amazon Redshift almacena datos DATE y TIMESTAMP de manera más eficiente que CHAR o VARCHAR, lo que da lugar a un mejor rendimiento en la consulta. Use el tipo de datos DATE o TIMESTAMP, según la resolución que necesite, en lugar de un tipo de carácter, al momento de almacenar información de fecha/hora. Para obtener más información, consulte Tipos de fecha y hora.

## Prácticas recomendadas de Amazon Redshift para la carga de datos

Cargar conjuntos de datos muy grandes puede tomar mucho tiempo y consumir gran cantidad de recursos de computación. La manera en que se carguen los datos también puede afectar el rendimiento de la consulta. En esta sección, se presentan las prácticas recomendadas para la carga eficiente de datos utilizando comandos COPY, inserciones masivas y tablas provisionales.

Temas

- Obtención de información sobre cómo cargar datos con un tutorial
- Uso del comando COPY para cargar datos
- Uso de un único comando COPY para cargar desde archivos múltiples
- Dividir los datos de carga
- Compresión de los archivos de datos
- Control de los archivos de datos antes y después de una carga
- Uso de una inserción de filas múltiples
- Uso de una inserción masiva
- Carga de datos en orden de clave de ordenación

- Carga de datos en bloques secuenciales
- Uso de tablas de series temporales
- Programación de periodos de mantenimiento

## Obtención de información sobre cómo cargar datos con un tutorial

En Tutorial: Carga de datos desde Amazon S3, se explica paso a paso el procedimiento completo para cargar datos a un bucket de Amazon S3 y usar después el comando COPY para cargar los datos en las tablas. En el curso, se incluye ayuda con la solución de errores de carga y se compara la diferencia de rendimiento entre la carga desde un archivo individual y la carga desde distintos archivos.

## Uso del comando COPY para cargar datos

El comando COPY carga los datos en paralelo desde Amazon S3, Amazon EMR, Amazon DynamoDB o distintos orígenes de datos en alojamientos remotos. COPY carga grandes cantidades de datos de manera mucho más eficiente que las instrucciones INSERT y también almacena los datos de manera más eficiente.

Para obtener más información acerca del uso del comando COPY, consulte Carga de datos desde Amazon S3 y Carga de datos desde una tabla de Amazon DynamoDB.

## Uso de un único comando COPY para cargar desde archivos múltiples

Amazon Redshift puede realizar cargas automáticas en paralelo desde varios archivos de datos comprimidos. Puede especificar los archivos que se cargarán mediante un prefijo de objeto de Amazon S3 o un archivo de manifiesto.

No obstante, si ejecuta varios comandos COPY a la vez para cargar una tabla desde varios archivos, Amazon Redshift se ve obligado a realizar una carga serializada. Este tipo de carga es mucho más lenta y requiere un proceso VACUUM al final si la tabla tiene una columna de ordenación definida. Para obtener más información acerca del comando COPY para cargar datos en paralelo, consulte Carga de datos desde Amazon S3.

## Dividir los datos de carga

Los archivos de origen de datos vienen en diferentes formatos y utilizan distintos algoritmos de compresión. Al cargar datos con el comando COPY, Amazon Redshift carga todos los archivos a los que hace referencia el prefijo del bucket de Amazon S3. (El prefijo es una cadena de caracteres al

principio del nombre de la clave de objeto). Si el prefijo hace referencia a varios archivos o archivos que se pueden dividir, Amazon Redshift carga los datos en paralelo, aprovechando la arquitectura MPP de Amazon Redshift. Esto divide la carga de trabajo entre los nodos del clúster. En contraste, cuando carga los datos desde un archivo que no se puede dividir, Amazon Redshift se ve obligado a realizar una carga serializada, que es mucho más lenta. En las siguientes secciones se describe la forma recomendada de cargar diferentes tipos de archivos en Amazon Redshift, en función del formato y la compresión.

### Carga de datos de archivos que se pueden dividir

Los siguientes archivos se pueden dividir automáticamente cuando se cargan los datos:

- un archivo CSV sin comprimir
- un archivo en columnas (Parquet/ORC)

Amazon Redshift divide automáticamente los archivos de 128 MB o más en fragmentos. Los archivos en columnas, específicamente Parquet y ORC, no se dividen si tienen menos de 128 MB. Redshift utiliza segmentos que funcionan en paralelo para cargar los datos. Esto proporciona un rendimiento de carga rápido.

### Carga de datos de archivos que no se pueden dividir

Los tipos de archivos como JSON o CSV, cuando se comprimen con otros algoritmos de compresión, como GZIP, no se dividen automáticamente. Para estos, le recomendamos que los divida manualmente en varios archivos más pequeños que tengan aproximadamente el mismo tamaño, de 1 MB a 1 GB después de la compresión. Además, la cantidad de archivos debe ser múltiplo de la cantidad de secciones en el clúster. Para obtener más información acerca de cómo dividir los datos en archivos múltiples y ejemplos de carga de datos con COPY, consulte Carga de datos desde Amazon S3.

## Compresión de los archivos de datos

Recomendamos que, cuando desee comprimir archivos de carga grandes, utilice gzip, lzop, bzip2 o Zstandard para comprimirlos y divida los datos en varios archivos más pequeños.

Especifique la opción GZIP, LZOP, BZIP2 o ZSTD con el comando COPY. En este ejemplo, se carga la tabla TIME desde un archivo lzop delimitado por la barra vertical.

```
copy time
```

```
from 's3://amzn-s3-demo-bucket/data/timerows.lzo'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
lzop
delimiter '|';
```

Hay instancias en las que no necesita dividir los archivos de datos sin comprimir. Para obtener más información sobre cómo dividir los datos y ejemplos de uso del comando COPY para cargar datos, consulte Carga de datos desde Amazon S3.

## Control de los archivos de datos antes y después de una carga

Antes de cargar los datos de Amazon S3, primero verifique que el bucket de Amazon S3 contiene todos los archivos correctos y solo esos. Para obtener más información, consulte Comprobación de que los archivos presentes en el bucket son los correctos.

Luego de que se complete la operación de carga, consulte la tabla de sistema STL_LOAD_COMMITS para controlar que se hayan cargado los archivos esperados. Para obtener más información, consulte Comprobación de carga correcta de datos.

## Uso de una inserción de filas múltiples

Si un comando COPY no es una opción y necesita inserciones SQL, utilice una inserción de múltiple filas, siempre que sea posible. La compresión de datos no es eficiente cuando agrega datos solamente a una fila o a pocas filas a la vez.

Las inserciones de filas múltiples mejoran el rendimiento al agrupar en lotes una serie de inserciones. En el siguiente ejemplo, se insertan tres filas en una tabla de cuatro columnas utilizando la instrucción INSERT. Se trata de una inserción pequeña que se muestra simplemente para ilustrar la sintaxis de una inserción de filas múltiples.

```
insert into category_stage values
(default, default, default, default),
(20, default, 'Country', default),
(21, 'Concerts', 'Rock', default);
```

Consulte INSERT para obtener más detalles y ejemplos.

## Uso de una inserción masiva

Use una operación de inserción masiva con una cláusula SELECT para una inserción de datos de alto rendimiento.

Utilice los comandos INSERT y CREATE TABLE AS cuando necesite mover datos o un subconjunto de datos de una tabla a otra.

Por ejemplo, la instrucción INSERT a continuación selecciona todas las filas de la tabla CATEGORY y las inserta en la tabla CATEGORY_STAGE.

```
insert into category_stage
(select * from category);
```

En el siguiente ejemplo, se crea CATEGORY_STAGE como una copia de CATEGORY y se insertan todas las filas de CATEGORY en CATEGORY_STAGE.

```
create table category_stage as
select * from category;
```

## Carga de datos en orden de clave de ordenación

Carga datos en orden de clave de ordenación para evitar la necesidad de limpieza.

Si cada lote de datos nuevos sigue las filas existentes en la tabla, sus datos se almacenarán correctamente en orden y no necesitará realizar una limpieza. No necesita ordenar previamente las filas en cada carga porque COPY ordena cada lote de datos entrantes a medida que se cargan.

Por ejemplo, imagine que carga datos todos los días en función de la actividad del día. Si la clave de ordenación es una columna de marca temporal, los datos se almacenan en orden. Este orden se produce porque los datos del día siempre se anexan después de los datos del día anterior. Para obtener más información, consulte Carga de los datos en orden de clave de clasificación. Para obtener más información sobre las operaciones de limpieza, consulte Limpieza de tablas.

## Carga de datos en bloques secuenciales

Si necesita agregar una gran cantidad de datos, cargue los datos en bloques secuenciales según el orden de ordenación para eliminar la necesidad de limpieza.

Por ejemplo, imagine que necesita cargar una tabla con eventos desde enero de 2017 hasta diciembre de 2017. Suponiendo que cada mes esté en un solo archivo, cargue las filas de enero, luego las de febrero y, así, sucesivamente. La tabla está completamente ordenada cuando se completa la carga y no es necesario ejecutar una limpieza. Para obtener más información, consulte Uso de tablas de series temporales.

Cuando cargue conjuntos muy grandes de datos, el espacio requerido para ordenarlos puede superar el espacio total disponible. Al cargar datos en bloques más pequeños, usa mucho menos espacio de ordenación intermedio durante cada carga. Además, la carga de bloques más pequeños facilita el reinicio si COPY genera un error y se revierte.

## Uso de tablas de series temporales

Si los datos tienen un periodo de retención fijo, puede organizarlos como una secuencia de tablas de series temporales. En esta secuencia, cada tabla es idéntica, pero contiene datos de diferentes intervalos de tiempo.

Puede eliminar fácilmente los datos antiguos; para ello, solo tiene que ejecutar el comando DROP TABLE en las tablas correspondientes. Este enfoque es mucho más rápido que ejecutar un proceso DELETE a gran escala y, además, le evita tener que ejecutar un proceso VACUUM posterior para recuperar espacio. Para ocultar el hecho de que los datos se almacenan en tablas diferentes, puede crear una vista UNION ALL. Cuando elimine los datos antiguos, ajuste la vista UNION ALL para eliminar las tablas descartadas. De igual modo, mientras carga periodos de tiempo nuevos en tablas nuevas, agregue las tablas nuevas a la vista. Para indicar al optimizador que omita el análisis de las tablas que no coincidan con el filtro de la consulta, la definición de vista filtra por el periodo de fechas correspondiente a cada tabla.

Procure no tener demasiadas tablas en la vista UNION ALL. Cada tabla adicional agrega tiempo de procesamiento a la consulta. No es necesario que las tablas utilicen la misma franja de tiempo. Puede tener tablas para diferentes periodos de tiempo (por ejemplo, diariamente, mensualmente y anualmente).

Si utiliza tablas de series temporales con una columna de marca temporal para la clave de ordenación, los datos se cargarán en orden de clave de ordenación. Así se elimina la necesidad de limpiar para volver a ordenar los datos. Para obtener más información, consulte Carga de los datos en orden de clave de clasificación.

## Programación de periodos de mantenimiento

Si se lleva a cabo un mantenimiento programado mientras se está ejecutando una consulta, esta se termina y se revierte y deberá reiniciarla. Programe operaciones de ejecución prolongada, como grandes cargas de datos o la operación VACUUM, para evitar realizarlas durante los periodos de mantenimiento. También puede reducir el riesgo y reiniciar de manera más fácil cuando lo necesite al realizar cargas de datos en incrementos más pequeños y administrar el tamaño de las operaciones

VACUUM. Para obtener más información, consulte Carga de datos en bloques secuenciales y Limpieza de tablas.

# Prácticas recomendadas de Amazon Redshift para el diseño de consultas

Para maximizar el rendimiento de las consultas, siga estas recomendaciones en el momento de crear las consultas.

- Diseñe las tablas según las prácticas recomendadas y, así, generará una base sólida para el rendimiento de las consultas. Para obtener más información, consulte Prácticas recomendadas de Amazon Redshift para el diseño de tablas.
- Evite usar `select *`. Incluya solamente las columnas que necesita.
- Utilice una expresión Expresión condicional CASE para realizar agregaciones complejas, en lugar de seleccionar de la misma tabla varias veces.
- No utilice combinaciones cruzadas a menos que sea absolutamente necesario. Estas combinaciones sin una condición de combinación dan lugar a un producto cartesiano de dos tablas. Por lo general, las combinaciones cruzadas se ejecutan como combinaciones de bucles anidados, que son los tipos de combinación más lentos.
- Utilice subconsultas en los casos donde una tabla en la consulta se utilice solamente para condiciones de predicado y la subconsulta devuelva una cantidad pequeña de filas (menos de 200). En el siguiente ejemplo, se utiliza una subconsulta para evitar combinar la tabla LISTING.

```
select sum(sales.qtysold)
from sales
where salesid in (select listid from listing where listtime > '2008-12-26');
```

- Utilice predicados para restringir el conjunto de datos tanto como sea posible.
- En el predicado, utilice los operadores menos costosos que pueda. Los operadores Condición de comparación son preferibles a los operadores LIKE. Se prefieren incluso los operadores LIKE a los SIMILAR TO o a los Operadores POSIX.
- Evite utilizar funciones en los predicados de consulta. Usarlos puede elevar el costo de la consulta al necesitar grandes cantidades de filas para resolver los pasos intermedios de la consulta.
- Si es posible, utilice una cláusula WHERE para restringir el conjunto de datos. El planificador de consultas puede utilizar el orden de las filas para determinar qué registros coinciden con los

criterios y, así, evitar examinar grandes cantidades de bloques de disco. Sin esto, el motor de ejecución de la consulta debe examinar en su totalidad todas las columnas que participan.

- Agregue predicados para filtrar tablas que participen en combinaciones, aun cuando se apliquen los mismos filtros. La consulta devuelve el mismo conjunto de resultados, pero Amazon Redshift puede filtrar las tablas combinadas antes del paso de análisis y, a continuación, puede omitir de manera eficiente el análisis de bloques de esas tablas. No se necesitan filtros redundantes si usted filtra una columna que se usa en la condición de combinación.

  Por ejemplo, imagine que desea combinar `SALES` y `LISTING` para encontrar ventas de tickets después de diciembre, agrupadas por vendedor. Ambas tablas se ordenan por fecha. La siguiente consulta combina las tablas según su clave común y filtra por valores `listing.listtime` mayores al 1. º de diciembre.

  ```
  select listing.sellerid, sum(sales.qtysold)
  from sales, listing
  where sales.salesid = listing.listid
  and listing.listtime > '2008-12-01'
  group by 1 order by 1;
  ```

  La cláusula WHERE no incluye un predicado para `sales.saletime`, por lo que el motor de ejecución debe examinar toda la tabla `SALES`. Si sabe que el filtro generaría que menos filas participen de la combinación, agregue también el filtro. En el siguiente ejemplo, se reduce considerablemente el tiempo de ejecución.

  ```
  select listing.sellerid, sum(sales.qtysold)
  from sales, listing
  where sales.salesid = listing.listid
  and listing.listtime > '2008-12-01'
  and sales.saletime > '2008-12-01'
  group by 1 order by 1;
  ```

- Utilice las claves de ordenación en la cláusula GROUP BY para que el planificador de consultas pueda utilizar la agregación de manera más eficiente. Una consulta puede calificar para agregación en una fase cuando la lista GROUP BY tiene solamente columnas de clave de ordenación, una de las cuales es también la clave de distribución. Las columnas con clave de ordenación en la lista GROUP BY deben incluir la primera clave de ordenación y, luego, las demás claves de ordenación que desee usar en el orden de la clave de ordenación. Por ejemplo, es válido utilizar la primera clave de ordenación, la primera y la segunda claves de ordenación, la primera, la

segunda y la tercera claves de ordenación y, así, sucesivamente. No es válido usar la primera y la tercera claves de ordenación.

Puede confirmar el uso de agregación en una fase al ejecutar el comando EXPLAIN y buscar `XN GroupAggregate` en el paso de agregación de la consulta.

- Si utiliza las cláusulas GROUP BY y ORDER BY, asegúrese de poner las columnas en el mismo orden en ambas cláusulas. Es decir, utilice el siguiente enfoque.

```
group by a, b, c
order by a, b, c
```

No utilice el siguiente enfoque.

```
group by b, c, a
order by a, b, c
```

# Seguimiento de las recomendaciones de Amazon Redshift Advisor

Para ayudarlo a mejorar el rendimiento y disminuir los costos operativos de su clúster de Amazon Redshift, Amazon Redshift Advisor le ofrece recomendaciones específicas sobre los cambios que deben realizarse. Advisor elabora sus recomendaciones personalizadas analizando el rendimiento y las métricas de uso de su clúster. Estas recomendaciones hechas a medida se relacionan con la configuración de las operaciones y el clúster. Para ayudarle a establecer una prioridad en las optimizaciones, Advisor clasifica las recomendaciones por orden de repercusión.

Advisor basa sus recomendaciones en observaciones sobre estadísticas de rendimiento o datos de operaciones. Advisor elabora las observaciones ejecutando pruebas en los clústeres para determinar si el valor de una prueba entra dentro de un rango especificado. Si el resultado de la prueba no entra dentro de dicho rango, Advisor genera una observación para el clúster. Al mismo tiempo Advisor crea una recomendación sobre cómo devolver el valor observado al rango de práctica recomendada. Advisor solo muestra recomendaciones que deberían tener una repercusión significativa sobre el rendimiento y las operaciones. Cuando Advisor concluye que una recomendación se ha aplicado, la elimina de la lista de recomendaciones.

Supongamos, por ejemplo, que su almacenamiento de datos contiene un gran número de columnas de tabla no comprimidas. En este caso, puede ahorrar costos en almacenamiento de clúster volviendo a crear tablas usando el parámetro `ENCODE` para especificar la compresión de columna.

En otro ejemplo, supongamos que Advisor observa que su clúster contiene una cantidad significativa de datos en datos de tabla sin comprimir. En este caso, le proporcionará el bloque de código SQL para encontrar las columnas de tabla que pueden comprimirse y los recursos que describen cómo comprimir dichas columnas.

## Regiones de Amazon Redshift en las que se admite Advisor

La característica Amazon Redshift Advisor solo está disponible en las siguientes regiones de AWS:

- Región EE. UU. Este (Norte de Virginia) (us-east-1)
- Región EE. UU. Este (Ohio) (us-east-2)
- Región EE. UU. Oeste (Norte de California) (us-west-1)
- Región EE. UU. Oeste (Oregón) (us-west-2)
- Región África (Ciudad del Cabo) (af-south-1)
- Región Asia-Pacífico (Hong Kong) (ap-east-1)
- Región Asia Pacífico (Hyderabad) (ap-south-2)
- Región Asia-Pacífico (Yakarta) (ap-southeast-3)
- Región Asia Pacífico (Melbourne) (ap-southeast-4)
- Región Asia-Pacífico (Malasia) (ap-southeast-5)
- Región Asia-Pacífico (Mumbai) (ap-south-1)
- Región Asia-Pacífico (Osaka) (ap-northeast-3)
- Región Asia-Pacífico (Seúl) (ap-northeast-2)
- Región Asia-Pacífico (Singapur) (ap-southeast-1)
- Región Asia-Pacífico (Sídney) (ap-southeast-2)
- Región Asia-Pacífico (Tokio) (ap-northeast-1)
- Región Canadá (Central) (ca-central-1)
- Región Oeste de Canadá (Calgary) (ca-west-1)
- Región China (Pekín) (cn-north-1)
- Región China (Ningxia) (cn-northwest-1)
- Región Europa (Fráncfort) (eu-central-1)
- Región Europa (Irlanda) (eu-west-1)
- Región Europa (Londres) (eu-west-2)

- Región Europa (Milán) (eu-south-1)
- Región Europa (París) (eu-west-3)
- Región Europa (España) (eu-south-2)
- Región Europa (Estocolmo) (eu-north-1)
- Región Europa (Zúrich) (eu-central-2)
- Región de Israel (Tel Aviv) (il-central-1)
- Región Medio Oriente (Baréin) (me-south-1)
- Región Medio Oriente (me-central-1)
- Región América del Sur (São Paulo) (sa-east-1)

Temas


- Visualización de las recomendaciones de Amazon Redshift Advisor
- Recomendaciones de Amazon Redshift Advisor


## Visualización de las recomendaciones de Amazon Redshift Advisor

Puede acceder a las recomendaciones de Amazon Redshift Advisor con la consola de Amazon Redshift, la API de Amazon Redshift o AWS CLI. Para acceder a las recomendaciones, debe tener un permiso `redshift:ListRecommendations` asociado al rol o identidad de IAM.

### Visualización de las recomendaciones de Amazon Redshift Advisor en la consola aprovisionada de Amazon Redshift

Puede consultar las recomendaciones de Amazon Redshift Advisor en la AWS Management Console.

Visualización de las recomendaciones de Amazon Redshift Advisor para los clústeres de Amazon Redshift en la consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. En el menú de navegación, elija Advisor.
3. Amplíe cada recomendación para ver más detalles. En esta página, puede ordenar y agrupar las recomendaciones.

### Visualización de las recomendaciones de Amazon Redshift Advisor con las operaciones de la API de Amazon Redshift

Puede mostrar las recomendaciones de Amazon Redshift Advisor para los clústeres de Amazon Redshift con la API de Amazon Redshift. Por lo general, debe desarrollar una aplicación en el lenguaje de programación que prefiera para llamar a la API `redshift:ListRecommendations` mediante un AWS SDK. Para obtener más información, consulte ListRecommendations en la Referencia de la API de Amazon Redshift.

### Visualización de las recomendaciones de Amazon Redshift Advisor con las operaciones de AWS Command Line Interface

Puede mostrar las recomendaciones de Amazon Redshift Advisor para los clústeres de Amazon Redshift con AWS Command Line Interface. Para obtener más información, consulte list-recommendations en la Referencia de comandos de la AWS CLI.

## Recomendaciones de Amazon Redshift Advisor

Amazon Redshift Advisor ofrece recomendaciones sobre cómo optimizar el clúster de Amazon Redshift para mejorar el rendimiento y ahorrarse costos operativos. Puede encontrar explicaciones sobre cada recomendación en la consola, tal y como se describe anteriormente. Encontrará más detalles sobre estas recomendaciones en las secciones siguientes.

Temas

- Comprimir objetos de archivos de Amazon S3 cargados con COPY
- Aislar varias bases de datos activas
- Reasignar memoria de administración de la carga de trabajo (WLM)
- Omitir el análisis de compresión durante la ejecución de COPY
- Dividir los objetos de Amazon S3 cargados con COPY
- Actualizar estadísticas de tablas
- Habilitar aceleración de consultas cortas
- Claves de distribución modificadas en tablas
- Modificar las claves de ordenación en tablas
- Modificar las codificaciones de compresión en columnas
- Recomendaciones de tipos de datos

## Comprimir objetos de archivos de Amazon S3 cargados con COPY

El comando COPY aprovecha la arquitectura del procesamiento masivo en paralelo (MPP) de Amazon Redshift para leer y cargar datos en paralelo. Puede leer archivos de Amazon S3, tablas de DynamoDB y texto generado en uno o más alojamientos remotos.

Cuando cargue grandes cantidades de datos, es muy recomendable ejecutar el comando COPY para cargar archivos de datos comprimidos desde S3. Comprimir grandes conjuntos de datos ahorra tiempo al cargar los archivos a Amazon S3. COPY también puede acelerar el proceso de carga al descomprimir los archivos a medida que se leen.

### Análisis

Los comandos COPY de ejecución prolongada que cargan grandes conjuntos de datos sin comprimir a menudo tienen la oportunidad de mejorar el rendimiento considerablemente. El análisis de Advisor identifica los comandos COPY que cargan grandes conjuntos de datos sin comprimir. En dicho caso Advisor genera una recomendación para comprimir los archivos de origen en Amazon S3.

### Recomendación

Asegúrese de que cada COPY que cargue una cantidad significativa de datos o se ejecute durante un periodo de tiempo significativo consuma objetos de datos comprimidos desde Amazon S3. Puede identificar los comandos COPY que cargan grandes conjuntos de datos sin comprimir desde Amazon S3 ejecutando el comando SQL siguiente como superusuario.

```
SELECT
    wq.userid, query, exec_start_time AS starttime, COUNT(*) num_files,
    ROUND(MAX(wq.total_exec_time/1000000.0),2) execution_secs,
    ROUND(SUM(transfer_size)/(1024.0*1024.0),2) total_mb,
    SUBSTRING(querytxt,1,60) copy_sql
FROM stl_s3client s
JOIN stl_query q USING (query)
JOIN stl_wlm_query wq USING (query)
WHERE s.userid>1 AND http_method = 'GET'
    AND POSITION('COPY ANALYZE' IN querytxt) = 0
    AND aborted = 0 AND final_state='Completed'
GROUP BY 1, 2, 3, 7
HAVING SUM(transfer_size) = SUM(data_size)
AND SUM(transfer_size)/(1024*1024) >= 5
ORDER BY 6 DESC, 5 DESC;
```

Si los datos transitorios permanecen en Amazon S3 después de cargarlos, lo que es habitual en las arquitecturas de lago de datos, almacenar estos datos de forma comprimida puede reducir los costos de almacenamiento.

Consejos de implementación

- El tamaño ideal de los objetos está comprendido entre 1 MB y 128 MB después de la compresión.
- Puede comprimir archivos con formato gzip, lzop o bzip2.

## Aislar varias bases de datos activas

Como práctica recomendada, es conveniente aislar entre sí las bases de datos en Amazon Redshift. Las consultas se ejecutan en una base de datos específica y no pueden obtener acceso a datos de otras bases de datos del clúster. No obstante, las consultas que ejecute en todas las bases de datos de un clúster comparten el mismo espacio de almacenamiento de clúster y los mismos recursos de computación subyacentes. Cuando un único clúster contiene varias bases de datos activas es habitual que sus cargas de trabajo no estén relacionadas.

Análisis

El análisis de Advisor revisa todas las bases de datos del clúster para detectar cargas de trabajo activas que se ejecutan al mismo tiempo. Si hay cargas de trabajo activas ejecutándose al mismo tiempo, Advisor genera una recomendación para considerar la migración de las bases de datos a clústeres independientes de Amazon Redshift.

Recomendación

Estudie la posibilidad de trasladar cada una de las bases de datos en las que se realizan consultas activas a un clúster exclusivo independiente. El uso de un clúster independiente puede reducir la contención de recursos y mejorar el rendimiento de las consultas. Esto se consigue porque permite establecer el tamaño de cada clúster en función de las necesidades de almacenamiento, costo y rendimiento de cada carga de trabajo. Asimismo, las cargas de trabajo no relacionadas a menudo se benefician de configuraciones de administración de cargas de trabajo diferentes.

Para identificar qué bases de datos se usan activamente puede ejecutar este comando de SQL como superusuario.

```
SELECT database,
  COUNT(*) as num_queries,
```

```
  AVG(DATEDIFF(sec,starttime,endtime)) avg_duration,
  MIN(starttime) as oldest_ts,
  MAX(endtime) as latest_ts
FROM stl_query
WHERE userid > 1
GROUP BY database;
```

Consejos de implementación

- Dado que un usuario tiene que conectarse con cada base de datos específicamente, y las consultas solo pueden obtener acceso a una única base de datos, el traslado de las bases de datos a clústeres independientes tiene repercusiones mínimas sobre los usuarios.
- Una opción para trasladar una base de datos consiste en seguir los pasos siguientes:
  1. Restaure temporalmente una instantánea del clúster actual en un clúster del mismo tamaño.
  2. Elimine todas las bases de datos del nuevo clúster salvo la base de datos de destino que va a trasladar.
  3. Cambie el tamaño del clúster a un tipo de nodo y un recuento apropiados para la carga de trabajo de la base de datos.

## Reasignar memoria de administración de la carga de trabajo (WLM)

Amazon Redshift dirige las consultas de los usuarios a Implementación de WLM manual para su procesamiento. La administración de cargas de trabajo (WLM) define el modo en que esas consultas se dirigen a las colas. Amazon Redshift asigna a cada cola una parte de la memoria disponible del clúster. La memoria de una cola se divide entre los slots de consultas de la cola.

Cuando se configura una cola con más ranuras de las necesarias para la carga de trabajo, la memoria asignada a dichas ranuras sin utilizar se infrautiliza. Si se reduce el número de ranuras configuradas para que se adapte a los requisitos de carga de trabajo punta, se redistribuirá la memoria infrautilizada a las ranuras activas y, como consecuencia, el rendimiento de las consultas mejora.

Análisis

El análisis de Advisor revisa los requisitos de simultaneidad de carga de trabajo para identificar las colas de consulta que tienen ranuras sin utilizar. Advisor genera una recomendación para reducir el número de ranuras de una cola cuando encuentra lo siguiente:

- Una cola con ranuras que están completamente inactivas durante todo el análisis.
- Una cola con más de cuatro ranuras con dos ranuras como mínimo inactivas durante todo el análisis.

### Recomendación

Si se reduce el número de ranuras configuradas para que se adapte a los requisitos de carga de trabajo punta, se redistribuirá la memoria infrautilizada a las ranuras activas. Considere reducir el número de ranuras configuradas en el caso de las colas en las que las ranuras nunca se han utilizado plenamente. Para identificar dichas colas puede comparar los requisitos de ranuras por hora máximos de cada cola ejecutando el siguiente comando de SQL como superusuario.

```
WITH
 generate_dt_series AS (select sysdate - (n * interval '5 second') as dt from (select
 row_number() over () as n from stl_scan limit 17280)),
 apex AS (
     SELECT iq.dt, iq.service_class, iq.num_query_tasks, count(iq.slot_count) as
 service_class_queries, sum(iq.slot_count) as service_class_slots
     FROM
         (select gds.dt, wq.service_class, wscc.num_query_tasks, wq.slot_count
         FROM stl_wlm_query wq
         JOIN stv_wlm_service_class_config wscc ON (wscc.service_class =
 wq.service_class AND wscc.service_class > 5)
         JOIN generate_dt_series gds ON (wq.service_class_start_time <= gds.dt AND
 wq.service_class_end_time > gds.dt)
         WHERE wq.userid > 1 AND wq.service_class > 5) iq
     GROUP BY iq.dt, iq.service_class, iq.num_query_tasks),
     maxes as (SELECT apex.service_class, trunc(apex.dt) as d, date_part(h,apex.dt) as
 dt_h, max(service_class_slots) max_service_class_slots
                     from apex group by apex.service_class, apex.dt,
 date_part(h,apex.dt))
SELECT apex.service_class - 5 AS queue, apex.service_class, apex.num_query_tasks AS
 max_wlm_concurrency, maxes.d AS day, maxes.dt_h || ':00 - ' || maxes.dt_h || ':59' as
 hour, MAX(apex.service_class_slots) as max_service_class_slots
FROM apex
JOIN maxes ON (apex.service_class = maxes.service_class AND apex.service_class_slots =
 maxes.max_service_class_slots)
GROUP BY  apex.service_class, apex.num_query_tasks, maxes.d, maxes.dt_h
ORDER BY apex.service_class, maxes.d, maxes.dt_h;
```

La columna `max_service_class_slots` representa la cantidad máxima de ranuras de consulta de WLM en la cola de consulta de dicha hora. Si hay colas infrautilizadas, implemente la optimización de reducción de slots mediante la modificación de un grupo de parámetros, tal como se describe en la Guía de administración de Amazon Redshift.

Consejos de implementación

- Si el volumen de su carga de trabajo es altamente variable, asegúrese de que el análisis capturó un periodo de utilización máxima. Si no capturó dicho periodo, ejecute el SQL anterior varias veces para supervisar los requisitos de simultaneidad máximos.
- Para obtener detalles acerca de cómo interpretar los resultados de la consulta del código SQL anterior, consulte wlm_apex_hourly.sql script en GitHub.

## Omitir el análisis de compresión durante la ejecución de COPY

Cuando carga datos en una tabla vacía con codificación de compresión que se declara con el comando COPY, Amazon Redshift aplica la compresión del almacenamiento. Esta optimización permite asegurarse de que los datos del clúster se almacenan de forma eficiente incluso cuando los cargan usuarios finales. El análisis necesario para aplicar una compresión puede requerir una cantidad significativa de tiempo.

Análisis

El análisis de Advisor busca si hay operaciones COPY que se hayan retrasado debido al análisis de compresión automático. El análisis determina las codificaciones de compresión muestreando los datos mientras estos se cargan. Este muestreo es similar al que ejecuta el comando ANALYZE COMPRESSION.

Cuando carga datos como parte de un proceso estructurado como, por ejemplo, un lote ETL (extracción, transformación, carga) de un día para el otro, puede definir previamente la compresión. También puede optimizar las definiciones de tabla para omitir esta fase permanentemente sin que tenga repercusiones negativas.

Recomendación

Para mejorar la capacidad de respuesta de COPY omitiendo la fase de análisis de compresión, implemente una de las dos opciones siguientes:

- Use el parámetro `ENCODE` de columna cuando cree tablas que cargue usando el comando COPY.

- Desactive la compresión suministrando el parámetro `COMPUPDATE OFF` en el comando COPY.

Normalmente la mejor solución consiste en usar la codificación de columnas durante la creación de la tabla, ya que este enfoque permite mantener el beneficio de almacenar los datos comprimidos en disco. Puede usar el comando ANALYZE COMPRESSION para sugerir codificaciones de compresión, pero tiene que volver a crear la tabla para aplicar estas codificaciones. Para automatizar este proceso, puede usar la utilidad AWSColumnEncodingUtility que se encuentra en GitHub.

Para identificar operaciones de COPY recientes que iniciaron el análisis de compresión automático, ejecute el comando de SQL siguiente.

```
  WITH xids AS (
    SELECT xid FROM stl_query WHERE userid>1 AND aborted=0
    AND querytxt = 'analyze compression phase 1' GROUP BY xid
    INTERSECT SELECT xid FROM stl_commit_stats WHERE node=-1)
SELECT a.userid, a.query, a.xid, a.starttime, b.complyze_sec,
    a.copy_sec, a.copy_sql
FROM (SELECT q.userid, q.query, q.xid, date_trunc('s',q.starttime)
    starttime, substring(querytxt,1,100) as copy_sql,
    ROUND(datediff(ms,starttime,endtime)::numeric / 1000.0, 2) copy_sec
    FROM stl_query q JOIN xids USING (xid)
    WHERE (querytxt ilike 'copy %from%' OR querytxt ilike '% copy %from%')
    AND querytxt not like 'COPY ANALYZE %') a
LEFT JOIN (SELECT xid,
    ROUND(sum(datediff(ms,starttime,endtime))::numeric / 1000.0,2) complyze_sec
    FROM stl_query q JOIN xids USING (xid)
    WHERE (querytxt like 'COPY ANALYZE %'
    OR querytxt like 'analyze compression phase %')
    GROUP BY xid ) b ON a.xid = b.xid
WHERE b.complyze_sec IS NOT NULL ORDER BY a.copy_sql, a.starttime;
```

Consejos de implementación

- Asegúrese de que todas las tablas de tamaño significativo creadas durante sus procesos de ETL (por ejemplo, tablas provisionales y tablas temporales) declaren una codificación de compresión para todas las columnas salvo para la primera clave de ordenación.
- Calcule el tamaño del tipo de vida previsto de la tabla que se está cargando para cada uno de los comandos COPY que el comando de SQL anterior identifica. Si está convencido de que la tabla

será muy pequeña, deshabilite la compresión junto con el parámetro `COMPUPDATE OFF`. O bien cree la tabla con una compresión explícita antes de cargarla con el comando COPY.

## Dividir los objetos de Amazon S3 cargados con COPY

El comando COPY aprovecha la arquitectura del procesamiento masivo en paralelo (MPP) de Amazon Redshift para leer y cargar datos de los archivos de Amazon S3. El comando COPY carga datos desde archivos múltiples en paralelo y divide, así, la carga de trabajo entre los nodos de su clúster. Para conseguir un rendimiento óptimo es muy recomendable que divida los datos en distintos archivos para aprovechar los beneficios del procesamiento en paralelo.

Análisis

El análisis de Advisor identifica los comandos COPY que cargan grandes conjuntos de datos contenidos en un pequeño número de archivos almacenados en Amazon S3. Los comandos COPY de ejecución prolongada que cargan grandes conjuntos de datos a partir de unos cuantos archivos a menudo tienen la oportunidad de mejorar el rendimiento considerablemente. Cuando Advisor detecta que estos comandos COPY necesitan una cantidad considerable de tiempo, crea una recomendación para aumentar el paralelismo dividiendo los datos en archivos adicionales en Amazon S3.

Recomendación

En este caso le recomendamos las siguientes acciones, enumeradas en orden de prioridad:

1. Optimice los comandos COPY que carguen menos archivos que el número de nodos de clúster.
2. Optimice los comandos COPY que carguen menos archivos que el número de sectores del clúster.
3. Optimice los comandos de COPY donde el número de archivos no sea un múltiplo del número de sectores del clúster.

Algunos comandos COPY cargan una cantidad de datos significativa o se ejecutan durante una cantidad de tiempo significativa. Para estos comandos, le recomendamos que cargue una cantidad de objetos de datos de Amazon S3 que sea equivalente a un múltiplo del número de sectores del clúster. Para identificar cuántos objetos S3 ha cargado cada comando COPY, ejecute el siguiente código SQL como superusuario.

```
SELECT
    query, COUNT(*) num_files,
    ROUND(MAX(wq.total_exec_time/1000000.0),2) execution_secs,
    ROUND(SUM(transfer_size)/(1024.0*1024.0),2) total_mb,
    SUBSTRING(querytxt,1,60) copy_sql
FROM stl_s3client s
JOIN stl_query q USING (query)
JOIN stl_wlm_query wq USING (query)
WHERE s.userid>1 AND http_method = 'GET'
    AND POSITION('COPY ANALYZE' IN querytxt) = 0
    AND aborted = 0 AND final_state='Completed'
GROUP BY query, querytxt
HAVING (SUM(transfer_size)/(1024*1024))/COUNT(*) >= 2
ORDER BY CASE
WHEN COUNT(*) < (SELECT max(node)+1 FROM stv_slices) THEN 1
WHEN COUNT(*) < (SELECT COUNT(*) FROM stv_slices WHERE node=0) THEN 2
ELSE 2+((COUNT(*) % (SELECT COUNT(*) FROM stv_slices))/(SELECT COUNT(*)::DECIMAL FROM
 stv_slices))
END, (SUM(transfer_size)/(1024.0*1024.0))/COUNT(*) DESC;
```

Consejos de implementación

- El número de sectores de un nodo depende del tamaño de nodo del clúster. Para obtener más información acerca de la cantidad de sectores en los diversos tipos de nodos, consulte Clústeres y nodos de Amazon Redshift en la Guía de administración de Amazon Redshift.
- Puede cargar varios archivos especificando un prefijo común o el prefijo de clave del conjunto, o enumerando explícitamente los archivos en un archivo de manifiesto. Para obtener más información acerca de cómo cargar archivos, consulte Carga de datos desde archivos comprimidos y sin comprimir.
- Amazon Redshift no tiene en cuenta el tamaño de los archivos a la hora de dividir la carga de trabajo. Divida los archivos de datos de carga de modo tal que sean de igual tamaño, entre 1 MB y 1 GB, después de la compresión.

## Actualizar estadísticas de tablas

Amazon Redshift usa un optimizador de consultas basado en el costo para elegir el mejor plan de ejecución de consultas. El cálculo de costos se basa en las estadísticas de tabla reunidas al ejecutar el comando ANALYZE. Cuando faltan estadísticas o están obsoletas, la base de datos puede elegir

un plan que sea menos eficiente para ejecutar la consulta, especialmente en el caso de las consultas complejas. Mantener las estadísticas actuales ayuda a que las consultas complejas se ejecuten en la menor cantidad de tiempo posible.

Análisis

El análisis de Advisor realiza el seguimiento de las tablas que carecen de estadísticas o cuyas estadísticas se han quedado obsoletas. Revisa los metadatos de acceso a la tabla asociados a consultas complejas. Si las tablas a las que se accede con frecuencia con patrones complejos carecen de estadísticas, Advisor crea una recomendación crítica para ejecutar ANALYZE. Si las tablas a las que se accede con frecuencia con patrones complejos tienen estadísticas obsoletas, Advisor crea una recomendación sugerida para ejecutar ANALYZE.

Recomendación

Cuando el contenido de una tabla cambia significativamente, actualice las estadísticas con ANALYZE. Recomendamos que ejecute ANALYZE siempre que se cargue un número significativo de filas de datos nuevas en una tabla ya existente con los comandos COPY o INSERT. También recomendamos ejecutar ANALYZE siempre que se modifique un número significativo de filas con los comandos UPDATE o DELETE. Para identificar las tablas que carecen de estadísticas o cuyas estadísticas están obsoletas, ejecute el siguiente comando de SQL como superusuario. Los resultados se ordenan de la tabla más grande a la más pequeña.

Para identificar las tablas que carecen de estadísticas o cuyas estadísticas están obsoletas, ejecute el siguiente comando de SQL como superusuario. Los resultados se ordenan de la tabla más grande a la más pequeña.

```
SELECT
   ti.schema||'.'||ti."table" tablename,
   ti.size table_size_mb,
   ti.stats_off statistics_accuracy
 FROM svv_table_info ti
 WHERE ti.stats_off > 5.00
 ORDER BY ti.size DESC;
```

Consejos de implementación

El umbral de ANALYZE predeterminado está establecido en 10 por ciento. Este valor predeterminado significa que el comando ANALYZE omite una tabla si menos del 10 por ciento de las filas de la tabla

han cambiado desde la última vez que se ejecutó este comando. Por lo tanto puede escoger ejecutar comandos ANALYZE al final de cada proceso ETL. Adoptar este enfoque significa que ANALYZE se omite con frecuencia, aunque también garantiza que se ejecutará cuando sea necesario.

Las estadísticas de ANALYZE tienen su máxima repercusión en las columnas que se usan en uniones (por ejemplo `JOIN tbl_a ON col_b`) o como predicados (por ejemplo `WHERE col_b = 'xyz'`). De forma predeterminada ANALYZE recopila estadísticas de todas las columnas de la tabla especificada. Si es preciso, puede reducir el tiempo necesario para ejecutar ANALYZE ejecutando ANALYZE solo en las columnas donde tenga más repercusión. Puede ejecutar el siguiente comando de SQL para identificar las columnas usadas como predicados. También puede dejar que Amazon Redshift elija las columnas que deben analizarse mediante la especificación de `ANALYZE PREDICATE COLUMNS`.

```
WITH predicate_column_info as (
SELECT ns.nspname AS schema_name, c.relname AS table_name, a.attnum as col_num,
 a.attname as col_name,
        CASE
            WHEN 10002 = s.stakind1 THEN array_to_string(stavalues1, '||')
            WHEN 10002 = s.stakind2 THEN array_to_string(stavalues2, '||')
            WHEN 10002 = s.stakind3 THEN array_to_string(stavalues3, '||')
            WHEN 10002 = s.stakind4 THEN array_to_string(stavalues4, '||')
            ELSE NULL::varchar
        END AS pred_ts
   FROM pg_statistic s
   JOIN pg_class c ON c.oid = s.starelid
   JOIN pg_namespace ns ON c.relnamespace = ns.oid
   JOIN pg_attribute a ON c.oid = a.attrelid AND a.attnum = s.staattnum)
SELECT schema_name, table_name, col_num, col_name,
       pred_ts NOT LIKE '2000-01-01%' AS is_predicate,
       CASE WHEN pred_ts NOT LIKE '2000-01-01%' THEN (split_part(pred_ts,
 '||',1))::timestamp ELSE NULL::timestamp END as first_predicate_use,
       CASE WHEN pred_ts NOT LIKE '%||2000-01-01%' THEN (split_part(pred_ts,
 '||',2))::timestamp ELSE NULL::timestamp END as last_analyze
FROM predicate_column_info;
```

Para obtener más información, consulte Análisis de tablas.

## Habilitar aceleración de consultas cortas

La aceleración de consultas cortas (SQA) da prioridad a una serie de consultas seleccionadas que se ejecutan rápidamente frente a consultas que tardan más en ejecutarse. SQA ejecuta las consultas

cortas en un espacio dedicado, de forma que estas consultas no tienen que esperar en las colas detrás de otras consultas más largas. SQA solamente da prioridad a las consultas de corta ejecución y a las consultas que están en una cola definida por el usuario. Con SQA, las consultas cortas se ejecutan con mayor rapidez y los usuarios tardan menos en ver los resultados.

Si activa SQA, puede reducir o eliminar las colas de administración de cargas de trabajo (WLM) dedicadas a las consultas cortas. Además, las consultas de larga ejecución no necesitan competir con las consultas cortas por los slots de las colas, por lo que las colas de WLM se pueden configurar para que utilicen menos slots de consulta. Si se utiliza una simultaneidad más baja, el rendimiento de las consultas aumenta y el rendimiento de todo el sistema mejora con la mayoría de las cargas de trabajo. Para obtener más información, consulte Aceleración de consultas cortas.

Análisis

Advisor comprueba los patrones de carga de trabajo y emite un informe con el número de consultas recientes sonde SQA debería reducir la latencia y el tiempo de cola diario para las consultas aptas para SQA.

Recomendación

Modifique la configuración de WLM para activar la aceleración de consultas cortas (SQA). Amazon Redshift utiliza un algoritmo de machine learning para analizar cada una de las consultas que reúnan los requisitos necesarios. Las predicciones mejoran, ya que SQA va aprendiendo de los patrones de consulta. Para obtener más información, consulte Configuración de la carga de trabajo.

Si se activa SQA, WLM establece en dynamic el tiempo de ejecución máximo de las consultas cortas de forma predeterminada. Le recomendamos que mantenga este valor para definir el tiempo de ejecución máximo de SQA.

Consejos de implementación

Para comprobar si la aceleración de consultas cortas (SQA) está activada, ejecute la consulta siguiente. Si la consulta devuelve una fila, SQA está habilitado.

```
select * from stv_wlm_service_class_config
where service_class = 14;
```

Para obtener más información, consulte Monitorización de SQA.

## Claves de distribución modificadas en tablas

Amazon Redshift distribuye las filas de las tablas en todo el clúster según el estilo de distribución de la tabla. Las tablas con distribución KEY requieren una columna como la clave de distribución (DISTKEY). Una fila de tablas se asigna a un sector del nodo de un clúster en función del valor de su columna DISTKEY.

Una DISTKEY adecuada coloca un número similar de filas en cada sector del nodo y es referenciado con frecuencia en las condiciones de combinación. Una combinación optimizada ocurre cuando las tablas se unen en las columnas DISTKEY, acelerando el rendimiento de la consulta.

Análisis

Advisor analiza la carga de trabajo de su clúster para identificar la clave de distribución más apropiada para las tablas que pueden beneficiarse de forma significativa de un estilo de distribución KEY.

Recomendación

Advisor da a ALTER TABLE instrucciones que alteran el DISTSTYLE y la DISTKEY de una tabla en función de su análisis. Para conseguir un beneficio de rendimiento significativo, asegúrese de que todas las instrucciones SQL dentro de un grupo de recomendaciones se implementan.

Redistribuir una tabla grande con ALTER TABLE consume recursos del clúster y requiere un bloqueo temporal de tablas en varios tiempos. Implemente cada grupo de recomendación cuando la carga de trabajo de los otros clústeres sea ligera. Puede encontrar más detalles sobre cómo optimizar las propiedades de distribución de las tablas en la publicación de blog Amazon Redshift Engineering's Advanced Table Design Playbook: Distribution Styles and Distribution Keys.

Para más información sobre ALTER DISTSTYLE y DISTKEY, vea ALTER TABLE.

> **Note**
>
> Si no ve una recomendación no significa necesariamente que los estilos de distribución actuales sean los más apropiados. Advisor no proporciona recomendaciones cuando no hay suficientes datos o el beneficio esperado de la redistribución es pequeño.
> Las recomendaciones de Advisor se aplican a una tabla particular y no se aplican necesariamente a una tabla que contiene una columna con el mismo nombre. Las tablas que comparten un nombre de columna pueden tener distintas características para dichas columnas, excepto si los datos contenidos en las tablas son los mismos.

Si ve recomendaciones para tablas de ensayo que son creadas o abandonadas por trabajos de ETL, modifique sus procesos ETL para utilizar las claves de distribución recomendadas de Advisor.

## Modificar las claves de ordenación en tablas

Amazon Redshift ordena las filas de la tabla según la clave de ordenación de la tabla. Las filas de la tabla se ordenan en función de los valores de columna de la clave de ordenación.

Ordenar una tabla por una clave de ordenación adecuada puede acelerar el rendimiento de las consultas, especialmente aquellas con predicados que tienen intervalos restringidos, al requerir que se lean menos bloques de tabla desde el disco.

Análisis

Advisor analiza la carga de trabajo del clúster durante varios días para identificar una clave de ordenación beneficiosa para las tablas.

Recomendación

Advisor proporciona dos grupos de instrucciones ALTER TABLE que modifican la clave de ordenación de una tabla en función de su análisis:

- Instrucciones que alteran una tabla que actualmente no tiene una clave de ordenación para agregar una clave de ordenación COMPOUND.
- Instrucciones que alteran una clave de ordenación de INTERLEAVED a COMPOUND o ninguna clave de ordenación.

  El uso de claves ordenación compuestas reduce de forma significativa la sobrecarga de mantenimiento. Las tablas con claves de ordenación compuestas no necesitan las caras operaciones VACUUM REINDEX que son necesarias para las ordenaciones intercaladas. En la práctica, las claves de ordenación compuestas son más efectivas que las claves de ordenación intercaladas para la gran mayoría de las cargas de trabajo de Amazon Redshift. No obstante, si una tabla es pequeña, es más eficiente no tener una clave de ordenación para evitar la sobrecarga del almacenamiento de claves de ordenación.

Al ordenar una tabla grande con ALTER TABLE, se consumen recursos de clúster y se requieren bloqueos de tabla en varios momentos. Implemente cada recomendación cuando la carga de trabajo

de un clúster sea moderada. Podrá encontrar más detalles sobre cómo optimizar las configuraciones de las claves de ordenación de las tablas en la publicación de blog Amazon Redshift Engineering's Advanced Table Design Playbook: Compound and Interleaved Sort Keys.

Para obtener más información acerca de ALTER SORTKEY, consulte ALTER TABLE.

> **Note**
>
> Si no ve una recomendación para una tabla, eso no significa necesariamente que la configuración actual sea la mejor. Advisor no proporciona recomendaciones cuando no hay suficientes datos o el beneficio esperado de la ordenación es pequeño.
> Las recomendaciones de Advisor se aplican a una tabla en particular y no se aplican necesariamente a una tabla que contiene una columna con el mismo nombre y tipo de datos. Las tablas que comparten nombres de columna pueden tener diferentes recomendaciones en función de los datos de las tablas y de la carga de trabajo.

## Modificar las codificaciones de compresión en columnas

La compresión es una operación que se produce en el nivel de las columnas que reduce el tamaño de los datos cuando se almacenan. La compresión se utiliza en Amazon Redshift para ahorrar espacio de almacenamiento y mejorar el rendimiento de las consultas mediante la reducción de la cantidad de operaciones E/S en el disco. Recomendamos una codificación de compresión óptima para cada columna en función de su tipo de datos y sus patrones de consulta. Con una compresión óptima, las consultas pueden ejecutarse de manera más eficiente, y la base de datos puede ocupar la menor cantidad de espacio de almacenamiento posible.

### Análisis

Advisor realiza de manera continua análisis de la carga de trabajo y el esquema de la base de datos del clúster para identificar la codificación de compresión óptima que corresponde a cada columna de la tabla.

### Recomendación

Advisor proporciona instrucciones ALTER TABLE que modifican la codificación de compresión de determinadas columnas en función de su análisis.

Cambiar las codificaciones de compresión de las columnas con ALTER TABLE consume recursos del clúster y requiere bloqueos en las tablas en distintos momentos. Es mejor implementar las recomendaciones cuando no es pesada la carga de trabajo del clúster.

Como referencia, Ejemplos de ALTER TABLE muestra varias instrucciones que cambian la codificación de una columna.

> **Note**
>
> Advisor no proporciona recomendaciones cuando no hay suficientes datos ni cuando es pequeño el beneficio que se espera del cambio en la codificación.

## Recomendaciones de tipos de datos

Amazon Redshift tiene una biblioteca de tipos de datos SQL para varios casos de uso. Estos incluyen tipos enteros como `INT` y tipos para almacenar personajes, como `VARCHAR`. Redshift almacena los tipos de forma optimizada para proporcionar un acceso rápido y un buen rendimiento de las consultas. Además, Redshift proporciona funciones para tipos específicos, que se pueden utilizar para dar formato o realizar cálculos en los resultados de las consultas.

### Análisis

Advisor realiza análisis de la carga de trabajo y el esquema de la base de datos del clúster de manera continua para identificar columnas que pueden beneficiarse considerablemente de un cambio de tipos de datos.

### Recomendación

Advisor proporciona una instrucción `ALTER TABLE` que agrega una nueva columna con el tipo de datos sugerido. Una instrucción `UPDATE` asociada copia los datos de la columna existente en la nueva columna. Después de crear la columna y cargar los datos, cambie las consultas y los scripts de ingesta para acceder a la nueva columna. A continuación, utilice las características y funciones especializadas en el nuevo tipo de datos, que se encuentra en Referencia de funciones SQL.

Copiar los datos existentes en la nueva columna puede llevar tiempo. Le recomendamos que implemente cada recomendación de Advisor cuando la carga de trabajo del clúster sea ligera. Consulte la lista de tipos de datos disponibles en Tipos de datos.

Tenga en cuenta que Advisor no proporciona recomendaciones cuando no hay suficientes datos ni cuando es pequeño el beneficio que se espera del cambio en el tipo de datos.

# Tutoriales de Amazon Redshift

Siga los pasos de estos tutoriales para obtener más información acerca de las características de Amazon Redshift:

| | |
|---|---|
| Tutorial: Carga de datos desde Amazon S3 | En este tutorial, siga el proceso de creación de datos en las tablas de bases de datos de Amazon Redshift desde archivos de datos en un bucket de S3 de principio a fin. |
| Tutorial: Consulta de datos anidados con Amazon Redshift Spectrum | En este tutorial, puede usar Redshift Spectrum para consultar los datos anidados. Redshift Spectrum le permite consultas datos en los formatos de archivo Parquet, ORC, JSON o Ion. |
| Tutorial: Configuración de colas de Administración de carga de trabajo (WLM) manual | En este tutorial, configurará Amazon Redshift para que utilice colas de administración de cargas de trabajo (WLM) manual. Amazon Redshift utiliza colas de WLM para administrar el modo en que se dividen los recursos para ejecutar consultas simultáneas. Debe configurar Amazon Redshift para que utilice WLM manual si necesita usar varias colas de WLM. |
| Tutorial: Uso de funciones SQL espaciales con Amazon Redshift | En este tutorial, consultará datos mediante funciones espaciales. Las funciones espaciales se utilizan para consultar datos geométricos y geográficos. |
| Tutoriales de Amazon Redshift ML | En estos tutoriales, creará y utilizará modelos de machine learning. |
| Tutorial: creación de roles y realización de consultas con RBAC | En este tutorial, utilizará el control de acceso basado en roles (RBAC) para crear y usar los permisos en una base de datos que ha |

| | |
|---|---|
| | creado. Con RBAC, puede crear roles con permisos específicos, como permisos de solo lectura o lectura y escritura. A continuación, podrá asignar estos roles a los usuarios para concederles los permisos especificados. |

# Optimización de tablas automática

La optimización automática de tablas es una capacidad de autoajuste que optimiza automáticamente el diseño de tablas mediante la aplicación de claves de ordenación y distribución sin necesidad de que intervenga el administrador. Mediante el uso de la automatización para ajustar el diseño de las tablas, puede empezar a trabajar y obtener el rendimiento más rápido sin invertir tiempo en ajustar e implementar manualmente las optimizaciones de las tablas.

La optimización automática de tablas observa continuamente la forma en que las consultas interactúan con las tablas. Utiliza métodos avanzados de inteligencia artificial que permiten elegir las claves de ordenación y distribución necesarias para optimizar el rendimiento de la carga de trabajo del clúster. Si Amazon Redshift determina que la aplicación de una clave mejora el rendimiento del clúster, las tablas se modifican de manera automática en cuestión de horas desde el momento en que se creó el clúster, con un impacto mínimo en las consultas.

Para aprovechar esta automatización, un administrador de Amazon Redshift crea una tabla nueva o modifica una ya existente para que pueda utilizar la optimización automática. Las tablas existentes con un estilo de distribución o una clave de ordenación de `AUTO` ya se encuentran habilitadas para la automatización. Cuando se ejecutan consultas en esas tablas, Amazon Redshift determina si una clave de ordenación o de distribución mejorará el rendimiento. En ese caso, Amazon Redshift modifica automáticamente la tabla sin necesidad de que intervenga el administrador. Si se ejecuta una cantidad mínima de consultas, las optimizaciones se aplican en cuestión de horas desde que se lanzó el clúster.

Si Amazon Redshift determina que una clave de distribución mejora el rendimiento de las consultas, las tablas en las que el estilo de distribución sea `AUTO` pueden cambiarlo a `KEY`.

Temas


- Habilitación, desactivación y supervisión de la optimización de tablas automática
- Compresión de columnas para reducir el tamaño de los datos almacenados
- Distribución de datos para la optimización de consultas
- Claves de clasificación
- Restricciones de tablas

# Habilitación, desactivación y supervisión de la optimización de tablas automática

De forma predeterminada, las tablas creadas sin definir explícitamente las claves de ordenación o las claves de distribución se establecen en AUTO. Asimismo, en el momento de crear la tabla, es posible establecer explícitamente una clave de ordenación o de distribución de forma manual. Si se establece la clave de ordenación o de distribución, la tabla no se administra automáticamente.

## Habilitación de la optimización automática de tablas

Si desea permitir que una tabla existente se optimice automáticamente, utilice las opciones de instrucción ALTER para cambiar la tabla a AUTO. Puede elegir definir la automatización para las claves de ordenación, pero no para las claves de distribución (y viceversa). Si se ejecuta una instrucción ALTER para convertir una tabla en una tabla automatizada, se conservan las claves de ordenación y los estilos de distribución existentes.

```
ALTER TABLE table_name ALTER SORTKEY AUTO;
```

```
ALTER TABLE table_name ALTER DISTSTYLE AUTO;
```

Para obtener más información, consulte ALTER TABLE.

En principio, una tabla no dispone de una clave de distribución ni una clave de ordenación. El estilo de distribución se establece en EVEN o ALL en función del tamaño de la tabla. A medida que se incrementa el tamaño de la tabla, Amazon Redshift aplica las claves de distribución y las claves de ordenación óptimas. Las optimizaciones se aplican en cuestión de horas una vez que se ejecuta una cantidad mínima de consultas. Cuando se determinan las optimizaciones de claves de ordenación, Amazon Redshift intenta optimizar los bloques de datos leídos desde el disco durante el análisis de una tabla. Cuando se determinan las optimizaciones de estilo de distribución, Amazon Redshift intenta optimizar el número de bytes que se transfieren entre los nodos del clúster.

## Eliminación de la optimización automática de tablas de una tabla

Puede eliminar una tabla de la optimización automática. La eliminación de una tabla de la automatización implica seleccionar una clave de ordenación o un estilo de distribución. Para cambiar el estilo de distribución, indique un estilo de distribución específico.

```
ALTER TABLE table_name ALTER DISTSTYLE EVEN;
```

```
ALTER TABLE table_name ALTER DISTSTYLE ALL;
```

```
ALTER TABLE table_name ALTER DISTSTYLE KEY DISTKEY c1;
```

Para cambiar una clave de ordenación, puede definir una clave de ordenación o no elegir ninguna.

```
ALTER TABLE table_name ALTER SORTKEY(c1, c2);
```

```
ALTER TABLE table_name ALTER SORTKEY NONE;
```

## Supervisión de la optimización de tablas automática

La vista del sistema SVV_ALTER_TABLE_RECOMMENDATIONS registra las recomendaciones actuales de Amazon Redshift Advisor para las tablas. En esta vista se muestran las recomendaciones para todas las tablas, tanto las que están definidas para la optimización automática como aquellas que no lo están.

Si desea ver si una tabla se ha definido para la optimización automática, consulte la vista del sistema SVV_TABLE_INFO. Las entradas aparecen únicamente para las tablas visibles en la base de datos de la sesión actual. Las recomendaciones se insertan en la vista dos veces al día a partir de las horas posteriores a la creación del clúster. Una vez que está disponible una recomendación, esta comienza en el plazo de una hora. Una vez aplicada una recomendación, sea por Amazon Redshift o por usted mismo, ya no aparecerá en la vista.

La vista del sistema SVL_AUTO_WORKER_ACTION muestra un registro de auditoría de todas las acciones que lleva a cabo Amazon Redshift y el estado anterior de la tabla.

La vista del sistema SVV_TABLE_INFO muestra una lista de todas las tablas del sistema, así como una columna donde se indica si la clave de ordenación y el estilo de distribución de la tabla están establecidos en AUTO.

Para obtener más información sobre cómo usar estas vistas del sistema, consulte Monitoreo del sistema (solo aprovisionado).

# Compresión de columnas para reducir el tamaño de los datos almacenados

La compresión es una operación en el nivel de columna que reduce el tamaño de los datos cuando se almacenan. La compresión mantiene el espacio de almacenamiento y reduce el tamaño de los datos que se leen desde el almacenamiento, lo que reduce la cantidad de E/S del disco y, por consiguiente, mejora el rendimiento de las consultas.

ENCODE AUTO es la opción predeterminada para las tablas. Cuando una tabla se establece a ENCODE AUTO, Amazon Redshift administra automáticamente la codificación de compresión para todas las columnas de la tabla. Para obtener más información, consulte CREATE TABLE y ALTER TABLE.

No obstante, si especifica la codificación de compresión para alguna columna de la tabla, esta ya no tendrá la opción ENCODE AUTO configurada. Amazon Redshift ya no administra automáticamente la codificación de compresión para todas las columnas de la tabla.

Al crear una tabla, se puede aplicar de forma manual un tipo de compresión, o codificación, a las columnas de esta. O bien, se puede utilizar el comando COPY para analizar y aplicar la compresión automáticamente. Para obtener más información, consulte Selección del comando COPY para las codificaciones de compresión. Para obtener información acerca de la aplicación de la compresión automática, consulte Carga de tablas con compresión automática.

> **Note**
>
> Le recomendamos enfáticamente que use el comando COPY para aplicar la compresión automática.

Puede elegir aplicar manualmente las codificaciones de compresión en caso de que la nueva tabla comparta las mismas características de datos que otra tabla. O puede hacerlo si detecta en las pruebas que las codificaciones de compresión aplicadas durante la compresión automática no son las más adecuadas para sus datos. Si selecciona aplicar codificaciones de compresión de forma manual, puede ejecutar el comando ANALYZE COMPRESSION en una tabla que ya tenga datos y utilizar los resultados para seleccionar las codificaciones de compresión.

Para aplicar la compresión de forma manual, debe especificar codificaciones de compresión para las columnas individuales como parte de la instrucción CREATE TABLE. La sintaxis es la siguiente.

```
CREATE TABLE table_name (column_name
data_type ENCODE encoding-type)[, ...]
```

En este caso, encoding-type se obtiene de la tabla de palabras clave de la siguiente sección.

Por ejemplo, la siguiente instrucción crea una tabla de dos columnas, la tabla PRODUCT. Cuando se cargan datos a la tabla, la columna PRODUCT_ID no se comprime, pero la columna PRODUCT_NAME sí, mediante una codificación por diccionario de bytes (BYTEDICT).

```
create table product(
product_id int encode raw,
product_name char(20) encode bytedict);
```

Puede especificar la codificación para una columna cuando se agrega a una tabla con el comando ALTER TABLE.

```
ALTER TABLE table-name ADD [ COLUMN ] column_name column_type ENCODE encoding-type
```

Temas

- Codificaciones de compresión
- Prueba de las codificaciones de compresión

## Codificaciones de compresión

Una codificación de compresión especifica el tipo de compresión que se aplica a una columna de valores de datos a medida que se agregan filas a una tabla.

ENCODE AUTO es la opción predeterminada para las tablas. Cuando una tabla se establece a ENCODE AUTO, Amazon Redshift administra automáticamente la codificación de compresión para todas las columnas de la tabla. Para obtener más información, consulte CREATE TABLE y ALTER TABLE.

No obstante, si especifica la codificación de compresión para alguna columna de la tabla, esta ya no tendrá la opción ENCODE AUTO configurada. Amazon Redshift ya no administra automáticamente la codificación de compresión para todas las columnas de la tabla.

Cuando utilice CREATE TABLE, ENCODE AUTO se deshabilita al especificar la codificación de compresión para cualquier columna de la tabla. Si se deshabilita ENCODE AUTO, Amazon Redshift

asigna de forma automática una codificación de compresión a las columnas para las que no se especifica un tipo ENCODE de la siguiente manera:

- A las columnas que están definidas como claves de ordenación se les asigna una compresión RAW.
- A las columnas que están definidas como tipos de datos BOOLEAN, REAL o DOUBLE PRECISION se les asigna una compresión RAW.
- Las columnas que se definen como tipos de datos SMALLINT, INTEGER, BIGINT, DECIMAL, CHAR, VARCHAR, DATE, TIMESTAMP o TIMESTAMPTZ tienen asignada la compresión AZ64.
- Las columnas que se definen como tipos de datos CHAR o VARCHAR tienen asignada la compresión LZO.

Puede cambiar la codificación de una tabla después de crearla mediante ALTER TABLE. Si deshabilita ENCODE AUTO mediante ALTER TABLE, Amazon Redshift dejará de administrar automáticamente las codificaciones de compresión de las columnas. Todas las columnas mantendrán los tipos de codificación de compresión que tenían cuando deshabilitó ENCODE AUTO hasta que los cambie o vuelva a habilitar ENCODE AUTO.

Amazon Redshift admite las siguientes codificaciones de compresión:

Raw

La codificación raw es la codificación predeterminada para las columnas designadas como claves de ordenación y para las columnas definidas como tipos de datos BOOLEAN, REAL o DOUBLE PRECISION. Con la codificación raw, los datos se almacenan descomprimidos y sin formato.

AZ64

AZ64 es un algoritmo codificado de compresión propiedad de Amazon que se ha diseñado para lograr una alta relación de compresión y un procesamiento mejorado de las consultas. En esencia, el algoritmo AZ64 comprime grupos de valores de datos más pequeños y utiliza instrucciones SIMD (una instrucción, varios datos) para realizar un procesamiento en paralelo. Utilice AZ64 para ahorrar una cantidad significativa de almacenamiento y conseguir un buen rendimiento con tipos de datos numéricos, de fecha y de hora.

Puede utilizar AZ64 como codificación de compresión para definir columnas con instrucciones CREATE TABLE y ALTER TABLE y los siguientes tipos de datos:

- SMALLINT

- INTEGER
- BIGINT
- DECIMAL
- DATE
- TIMESTAMP
- TIMESTAMPTZ

### Byte-dictionary

En la codificación por diccionarios de bytes, se crea un diccionario independiente de valores únicos para cada bloque de los valores de columna del disco. (Un bloque de disco de Amazon Redshift ocupa 1 MB). El diccionario tiene hasta 256 valores de un byte que se almacenan como índices de los valores de datos originales. Si se almacenan más de 256 valores en un mismo bloque, los valores adicionales se graban en un bloque descomprimido sin formato. El proceso se repite para cada bloque del disco.

Esta codificación es muy eficaz en columnas de cadenas de cardinalidad baja. Esta codificación es una solución óptima cuando el dominio de los datos de una columna es menor que 256 valores únicos.

En las columnas cuyo tipo de datos de cadena (CHAR y VARCHAR) está codificado con BYTEDICT, Amazon Redshift realiza exámenes vectorizados y evaluaciones de predicados que actúan directamente en los datos comprimidos. Estos exámenes utilizan instrucciones de instrucción única y datos múltiples (SIMD) específicas del hardware para el procesamiento paralelo. Esto acelera considerablemente el examen de las columnas de cadenas. La codificación por diccionario de bytes es muy eficaz en cuanto a espacio si una columna CHAR/VARCHAR almacena cadenas largas de caracteres.

Supongamos que una tabla tiene una columna COUNTRY con un tipo de datos CHAR(30). A medida que se cargan datos, Amazon Redshift crea un diccionario y rellena la columna COUNTRY con el valor índice. El diccionario tiene los valores únicos indexados y la tabla en sí tiene solo los subscripts de un byte de los valores correspondientes.

> **Note**
>
> Los espacios a la derecha se almacenan en columnas de caracteres de longitud fija. Por lo tanto, en una columna CHAR (30), cada valor comprimido ahorra 29 bytes de almacenamiento cuando utiliza la codificación por diccionario de bytes.

En la siguiente tabla, se representa el diccionario para la columna COUNTRY.

| Valor único de datos | Índice en diccionario | Tamaño (longitud fija, 30 bytes por valor) |
|---|---|---|
| England | 0 | 30 |
| United States of America | 1 | 30 |
| Venezuela | 2 | 30 |
| Sri Lanka | 3 | 30 |
| Argentina | 4 | 30 |
| Japan | 5 | 30 |
| Total | | 180 |

En la siguiente tabla, se representan los valores de la columna COUNTRY.

| Valor original del dato | Tamaño original (longitud fija, 30 bytes por valor) | Valor comprimido (índice) | Tamaño nuevo (bytes) |
|---|---|---|---|
| England | 30 | 0 | 1 |
| England | 30 | 0 | 1 |
| United States of America | 30 | 1 | 1 |

| Valor original del dato | Tamaño original (longitud fija, 30 bytes por valor) | Valor comprimido (índice) | Tamaño nuevo (bytes) |
|---|---|---|---|
| United States of America | 30 | 1 | 1 |
| Venezuela | 30 | 2. | 1 |
| Sri Lanka | 30 | 3 | 1 |
| Argentina | 30 | 4 | 1 |
| Japan | 30 | 5 | 1 |
| Sri Lanka | 30 | 3 | 1 |
| Argentina | 30 | 4 | 1 |
| Total | 300 | | 10 |

El tamaño total comprimido en este ejemplo se calcula de la siguiente manera: hay 6 entradas diferentes almacenadas en el diccionario (6 * 30 = 180) y la tabla tiene 10 valores comprimidos de 1 byte para un total de 190 bytes.

Delta

Las codificaciones Delta son de gran utilidad para las columnas de fecha y hora.

La codificación Delta comprime los datos al registrar la diferencia entre los valores que se suceden en la columna. Esta diferencia se registra en un diccionario independiente para cada bloque de valores de columnas del disco. (Un bloque de disco de Amazon Redshift ocupa 1 MB). Por ejemplo, supongamos que la columna contiene 10 valores enteros en secuencia del 1 al 10. El primero se almacena como un valor entero de 4 bytes (más una marca de 1 byte). Los nueve siguientes se almacenan cada uno como un byte con el valor 1, lo que indica que es un valor mayor que el anterior.

La codificación Delta tiene dos variaciones:

- DELTA registra las diferencias como valores de 1 byte (enteros de 8 bits)

- DELTA32K registra las diferencias como valores de 2 bytes (enteros de 16 bits)

Si la mayoría de los valores de la columna se pudieran comprimir con un solo byte, la variación de 1 byte resultaría muy eficaz. No obstante, si los deltas son mayores, esta codificación, en el peor de los casos, es de alguna manera menos eficaz que el almacenamiento de los datos sin comprimir. Se aplica una lógica similar a la versión de 16 bits.

Si la diferencia entre dos valores supera el rango de 1 byte (DELTA) o el rango de 2 bytes (DELTA32K), se almacena el valor original completo, con una marca de 1 byte al principio. El rango de 1 byte abarca desde -127 hasta 127 y el rango de 2 bytes desde -32K hasta 32K.

En la siguiente tabla, se muestra cómo funciona una codificación delta para una columna numérica.

| Valor original del dato | Tamaño original (bytes) | Diferencia (delta) | Valor comprimido | Tamaño comprimido (bytes) |
|---|---|---|---|---|
| 1 | 4 | | 1 | 1 + 4 (marca + valor real) |
| 5 | 4 | 4 | 4 | 1 |
| 50 | 4 | 45 | 45 | 1 |
| 200 | 4 | 150 | 150 | 1 + 4 (marca + valor real) |
| 185 | 4 | -15 | -15 | 1 |
| 220 | 4 | 35 | 35 | 1 |
| 221 | 4 | 1 | 1 | 1 |
| Totales | 28 | | | 15 |

LZO

La codificación LZO proporciona una relación de compresión muy alta con un buen rendimiento. La codificación LZO funciona especialmente bien para las columnas CHAR y VARCHAR que almacenan cadenas de caracteres muy largas. Resulta especialmente adecuada para textos de formato libre, como descripciones de productos, comentarios de usuarios o cadenas JSON.

Mostly

Las codificaciones mostly son útiles cuando el tipo de datos de una columna es mayor que lo que requieren la mayoría de los valores almacenados. Al especificar una codificación mostly para este tipo de columnas, puede comprimir la mayoría de los valores de la columna a un tamaño de almacenamiento estándar más pequeño. Los valores restantes que no pueden comprimirse se almacenan sin formato. Por ejemplo, puede comprimir una columna de 16 bits, como una columna IT2, a un almacenamiento de 8 bits.

Por lo general, las codificaciones mostly funcionan con los siguientes tipos de datos:

- SMALLINT/INT2 (16 bits)
- INTEGER/INT (32 bits)
- BIGINT/INT8 (64 bits)
- DECIMAL/NUMERIC (64 bits)

Seleccione la variación adecuada de codificación mostly en función del tamaño de los tipos de datos de la columna. Por ejemplo, aplique la codificación MOSTLY8 a una columna definida para valores enteros de 16 bits. No está permitido aplicar MOSTLY16 a una columna con tipos de datos de 16 bits o MOSTLY32 a una columna con tipos de datos de 32 bits.

La mayoría de las codificaciones pueden ser menos eficaces que la ausencia de compresión cuando no se puede comprimir un número relativamente elevado de los valores de la columna. Antes de aplicar una de estas codificaciones a una columna, realice la comprobación correspondiente. La mayoría de los valores que se vayan a cargar ahora (y que es probable que se carguen más adelante) deberían ajustarse a los rangos que se muestran en la siguiente tabla.

| Codificación | Tamaño de almacenamiento comprimido | Rango de valores que pueden comprimirse (los valores fuera del rango se almacenan sin formato) |
|---|---|---|
| MOSTLY8 | 1 byte (8 bits) | De -128 a 127 |
| MOSTLY16 | 2 bytes (16 bits) | De -32768 a 32767 |
| MOSTLY32 | 4 bytes (32 bits) | De -2147483648 a 2147483647 |

> **Note**
>
> Para los valores decimales, omita la coma decimal para determinar si el valor se ajusta al rango. Por ejemplo, 1 234,56 es tratado como 123 456 y se puede comprimir en una columna MOSTLY32.

Por ejemplo, la columna VENUEID de la tabla VENUE está definida como una columna de valores enteros sin formato, lo que significa que su valor consume 4 bytes de almacenamiento. No obstante, el rango actual de valores de la columna es de **`0`** a **`309`**. Por lo tanto, si se vuelve a crear y a cargar esta tabla con la codificación MOSTLY16 para VENUEID, se reduciría el almacenamiento de cada valor en esa columna a 2 bytes.

Si la mayoría de los valores de VENUEID a los que se hace referencia en otra tabla estuvieran en el rango de 0 a 127, podría tener sentido codificar esa columna de clave externa como MOSTLY8. Antes de tomar una decisión, ejecute algunas consultas en los datos de la tabla de referencia para averiguar si la mayoría de los valores están comprendidos en el intervalo de 8 bits, 16 bits o 32 bits.

En la siguiente tabla, se muestran los tamaños comprimidos para valores numéricos específicos cuando se utilizan codificaciones MOSTLY8, MOSTLY16 y MOSTLY32:

| Valor original | Tamaño original INT o BIGINT (bytes) | Tamaño comprimido con MOSTLY8 (bytes) | Tamaño comprimido con MOSTLY16 (bytes) | Tamaño comprimido con MOSTLY32 (bytes) |
|---|---|---|---|---|
| 1 | 4 | 1 | 2 | 4 |
| 10 | 4 | 1 | 2 | 4 |
| 100 | 4 | 1 | 2 | 4 |
| 1 000 | 4 | Igual que el tamaño de datos sin formato | 2 | 4 |
| 10000 | 4 | | 2 | 4 |
| 20000 | 4 | | 2 | 4 |
| 40000 | 8 | | Igual que el tamaño de datos sin formato | 4 |
| 100000 | 8 | | | 4 |
| 2000000000 | 8 | | | 4 |

Run length

La codificación de run length reemplaza un valor que se repite de manera consecutiva por un token que consiste en el valor y un recuento de la cantidad de ocurrencias consecutivas (la longitud de la ejecución). Se crea un diccionario independiente de valores únicos para cada bloque de los valores de columna del disco. (Un bloque de disco de Amazon Redshift ocupa 1 MB). Esta codificación es una opción ideal para una tabla en la que los valores de datos suelen repetirse de manera consecutiva; por ejemplo, cuando la tabla está ordenada según esos valores.

Por ejemplo, imagine que una columna de una tabla de grandes dimensiones presenta un dominio previsiblemente pequeño, como una columna COLOR con menos de 10 valores posibles. Es probable que estos valores queden dispuestos en secuencias de gran longitud en toda la tabla, incluso si los datos no están ordenados.

No recomendamos aplicar la codificación de run length en ninguna columna que esté designada como clave de ordenación. Los exámenes de rango restringido funcionan mejor cuando los bloques tienen una cantidad similar de filas. Si las columnas de clave de ordenación se comprimen mucho más que otras columnas en la misma consulta, los exámenes de rango restringido podrían tener un rendimiento deficiente.

En la siguiente tabla se utiliza el ejemplo de la columna COLOR para mostrar cómo funciona la codificación de run length.

| Valor original del dato | Tamaño original (bytes) | Valor comprimido (token) | Tamaño comprimido (bytes) |
|---|---|---|---|
| Blue | 4 | {2,Blue} | 5 |
| Blue | 4 | | 0 |
| Green | 5 | {3,Green} | 6 |
| Green | 5 | | 0 |
| Green | 5 | | 0 |
| Blue | 4 | {1,Blue} | 5 |
| Yellow | 6 | {4,Yellow} | 7 |
| Yellow | 6 | | 0 |
| Yellow | 6 | | 0 |
| Yellow | 6 | | 0 |
| Total | 51 | | 23 |

### Text255 and Text32k

Las codificaciones text255 y text32k son útiles para comprimir columnas VARCHAR en las que se repiten con frecuencia las mismas palabras. Se crea un diccionario independiente de palabras únicas para cada bloque de los valores de columna del disco. (Un bloque de disco de Amazon Redshift ocupa 1 MB). El diccionario tiene las primeras 245 palabras únicas de la columna. Estas

palabras se reemplazan en el disco por un valor índice de un byte, lo que representa uno de 245 valores, y todas las palabras que no están representadas en el diccionario se almacenan sin comprimir. El proceso se repite para cada bloque de 1 MB del disco. Si las palabras indexadas se repiten con frecuencia en la columna, esta arrojará una alta relación de compresión.

Para la codificación text32k, se aplica el mismo principio, pero el diccionario para cada bloque no captura una cantidad específica de palabras. En lugar de ello, el diccionario indexa cada palabra única que encuentra hasta que las entradas combinadas alcancen una longitud de 32K, menos alguna sobrecarga. Los valores índices se almacenan en dos bytes.

Por ejemplo, veamos el caso de la columna VENUENAME en la tabla VENUE. Palabras como **`Arena`**, **`Center`** y **`Theatre`** son recurrentes en esta columna y es posible que estén dentro de las primeras 245 palabras que se encuentran en cada bloque si se aplica la compresión text255. En tal caso, esta columna se beneficia de la compresión. El motivo de ello es que cada vez que aparezcan esas palabras, ocuparán solo 1 byte de almacenamiento (en lugar de 5, 6 o 7 bytes, respectivamente).

ZSTD

La codificación zstandard (ZSTD) proporciona una relación de compresión muy alta con un muy buen rendimiento en diferentes conjuntos de datos. La codificación ZSTD funciona especialmente bien para las columnas CHAR y VARCHAR que almacenan una gran variedad de cadenas de caracteres largas y cortas, como descripciones de productos, comentarios de usuarios, registros o cadenas JSON. Mientras que algunos algoritmos, como la codificación Delta o la codificación Mostly pueden, potencialmente, usar más espacio de almacenamiento que si se dejan los datos sin comprimir, es poco probable que ZSTD aumente el uso del disco.

ZSTD admite los tipos de datos SMALLINT, INTEGER, BIGINT, DECIMAL, REAL, DOUBLE PRECISION, BOOLEAN, CHAR, VARCHAR, DATE, TIMESTAMP y TIMESTAMPTZ.

En la siguiente tabla, se identifican las codificaciones de compresión compatibles y los tipos de datos que admiten codificación.

| Tipo de codificación | Palabra clave en CREATE TABLE y ALTER TABLE | Tipos de datos |
| --- | --- | --- |
| Raw (sin comprimir) | RAW | Todos |

| Tipo de codificación | Palabra clave en CREATE TABLE y ALTER TABLE | Tipos de datos |
|---|---|---|
| AZ64 | AZ64 | SMALLINT, INTEGER, BIGINT, DECIMAL, DATE, TIMESTAMP, TIMESTAMPTZ |
| Diccionario de bytes | BYTEDICT | SMALLINT, INTEGER, BIGINT, DECIMAL, REAL, DOUBLE PRECISION, CHAR, VARCHAR, DATE, TIMESTAMP, TIMESTAMP TZ |
| Delta | DELTA<br>DELTA32K | SMALLINT, INT, BIGINT, DATE, TIMESTAMP, DECIMAL<br>INT, BIGINT, DATE, TIMESTAMP, DECIMAL |
| LZO | LZO | SMALLINT, INTEGER, BIGINT, DECIMAL, CHAR, VARCHAR, DATE, TIMESTAMP, TIMESTAMP TZ, SUPER |
| Mostlyn | MOSTLY8<br>MOSTLY16<br>MOSTLY32 | SMALLINT, INT, BIGINT, DECIMAL<br>INT, BIGINT, DECIMAL<br>BIGINT, DECIMAL |
| Run-length | RUNLENGTH | SMALLINT, INTEGER, BIGINT, DECIMAL, REAL, DOUBLE PRECISION, BOOLEAN, CHAR, VARCHAR, DATE, TIMESTAMP, TIMESTAMPTZ |
| Texto | TEXT255<br>TEXT32K | Solo VARCHAR<br>Solo VARCHAR |

| Tipo de codificación | Palabra clave en CREATE TABLE y ALTER TABLE | Tipos de datos |
|---|---|---|
| Zstandard | ZSTD | SMALLINT, INTEGER, BIGINT, DECIMAL, REAL, DOUBLE PRECISION, BOOLEAN, CHAR, VARCHAR, DATE, TIMESTAMP, TIMESTAMPTZ, SUPER |

## Prueba de las codificaciones de compresión

Si decide especificar de forma manual las codificaciones de las columnas, posiblemente quiera probar los diferentes tipos de compresión con sus datos.

> **Note**
>
> Le recomendamos utilizar el comando COPY para cargar los datos siempre que sea posible y permitir que el comando COPY seleccione las codificaciones óptimas según sus datos. También puede utilizar el comando ANALYZE COMPRESSION para ver las codificaciones sugeridas para los datos existentes. Para obtener información acerca de la aplicación de la compresión automática, consulte Carga de tablas con compresión automática.

Para que la prueba de compresión de datos sea significativa, debe tener una gran cantidad de filas. Para este ejemplo, se crea una tabla y se insertan filas mediante el uso de una instrucción que selecciona datos de dos tablas, VENUE y LISTING. Se omite la cláusula WHERE que, por lo general, uniría las dos tablas. Como resultado de esto, cada fila de la tabla VENUE se une a todas las filas de la tabla LISTING, lo que supone un total de más de 32 millones de filas. Esta operación se la conoce como unión cartesiana y, por lo general, no se la recomienda. No obstante, a estos efectos, se trata de un método cómodo para crear muchas filas. Puede omitir este paso si ya tiene una tabla existente con datos que desea probar.

Después de tener una tabla con datos de muestra, se crea una tabla con siete columnas. Cada una de ellas cuenta con una codificación de compresión diferente: raw, bytedict, lzo, run length, text255, text32k y zstd. Se rellena cada columna con los mismos datos por medio de la ejecución de un comando INSERT que selecciona los datos de la primera tabla.

Si desea probar las codificaciones de compresión, haga lo siguiente:

1. (Opcional) Primero, utilice una unión cartesiana para crear una tabla con una gran cantidad de filas. Omita este paso si desea realizar la prueba con una tabla existente.

```
create table cartesian_venue(
venueid smallint not null distkey sortkey,
venuename varchar(100),
venuecity varchar(30),
venuestate char(2),
venueseats integer);

insert into cartesian_venue
select venueid, venuename, venuecity, venuestate, venueseats
from venue, listing;
```

2. Luego, cree una tabla con las codificaciones que desea comparar.

```
create table encodingvenue (
venueraw varchar(100) encode raw,
venuebytedict varchar(100) encode bytedict,
venuelzo varchar(100) encode lzo,
venuerunlength varchar(100) encode runlength,
venuetext255 varchar(100) encode text255,
venuetext32k varchar(100) encode text32k,
venuezstd varchar(100) encode zstd);
```

3. Inserte los mismos datos en todas las columnas utilizando una instrucción INSERT con una cláusula SELECT.

```
insert into encodingvenue
select venuename as venueraw, venuename as venuebytedict, venuename as venuelzo,
 venuename as venuerunlength, venuename as  venuetext32k, venuename as  venuetext255,
 venuename as venuezstd
from cartesian_venue;
```

4. Compruebe la cantidad de filas en la tabla nueva.

```
select count(*) from encodingvenue

  count
----------
```

```
 38884394
(1 row)
```

5. Consulte la tabla de sistema STV_BLOCKLIST para comparar la cantidad de bloques de 1 MB del disco que usa cada columna.

   La función de agregación MAX devuelve la cantidad máxima de bloques para cada columna. La tabla STV_BLOCKLIST incluye detalles de las tres columnas generadas por el sistema. En este ejemplo se utiliza `col < 6` en la cláusula WHERE para excluir las columnas generadas por el sistema.

```
select col, max(blocknum)
from stv_blocklist b, stv_tbl_perm p
where (b.tbl=p.id) and name ='encodingvenue'
and col < 7
group by name, col
order by col;
```

   La consulta devuelve los siguientes resultados. Las columnas están numeradas a partir del cero. Los resultados pueden tener números diferentes, en función de cómo esté configurado su clúster, pero los tamaños relativos deben ser similares. Puede observar que la codificación BYTEDICT en la segunda columna produjo los mejores resultados para este conjunto de datos. Este enfoque tiene una relación de compresión de más de 20:1. La codificación LZO y ZSTD también produjeron resultados excelentes. Desde luego, los diferentes conjuntos de datos producen resultados distintos. Cuando una columna tiene cadenas de texto más largas, LZO suele producir los mejores resultados de compresión.

```
 col | max
-----+-----
   0 | 203
   1 |  10
   2 |  22
   3 | 204
   4 |  56
   5 |  72
   6 |  20
(7 rows)
```

Si tiene datos en una tabla existente, puede utilizar el comando ANALYZE COMPRESSION para ver las codificaciones sugeridas para la tabla. Por ejemplo, en el siguiente ejemplo, se muestra la codificación recomendada para una copia de la tabla VENUE, CARTESIAN_VENUE, que tiene 38 millones de filas. Observe que ANALYZE COMPRESSION recomienda utilizar la codificación LZO para la columna VENUENAME. ANALYZE COMPRESSION selecciona la compresión óptima en función de varios factores, incluido el porcentaje de reducción. En este caso en particular, BYTEDICT proporciona una mejor compresión, pero LZO también produce una compresión mayor que el 90 por ciento.

```
analyze compression cartesian_venue;

Table          | Column     | Encoding | Est_reduction_pct
---------------+------------+----------+------------------
reallybigvenue | venueid    | lzo      | 97.54
reallybigvenue | venuename  | lzo      | 91.71
reallybigvenue | venuecity  | lzo      | 96.01
reallybigvenue | venuestate | lzo      | 97.68
reallybigvenue | venueseats | lzo      | 98.21
```

## Ejemplo

El ejemplo siguiente crea una tabla CUSTOMER que tiene columnas con diferentes tipos de datos. Esta instrucción CREATE TABLE muestra una de las combinaciones posibles de codificaciones de compresión para estas columnas.

```
create table customer(
custkey int encode delta,
custname varchar(30) encode raw,
gender varchar(7) encode text255,
address varchar(200) encode text255,
city varchar(30) encode text255,
state char(2) encode raw,
zipcode char(5) encode bytedict,
start_date date encode delta32k);
```

En la siguiente tabla, se muestran las codificaciones de columnas que se eligieron para la tabla CUSTOMER y se da una explicación de las elecciones:

| Columna | Tipo de datos: | Codificación | Explicación |
|---|---|---|---|
| CUSTKEY | int | delta | CUSTKEY consta de valores enteros únicos consecutivos. Como las diferenci as son de un byte, DELTA es una buena opción. |
| CUSTNAME | varchar (30) | raw | CUSTNAME tiene un dominio mayor con menos valores repetidos. Es posible que cualquier codificación de compresión sea ineficaz. |
| GENDER | varchar (7) | text255 | GENDER es un dominio muy pequeño con mucho valores repetidos. La codificación text255 funciona bien para las columnas VARCHAR, en las que se repiten las mismas palabras. |
| ADDRESS | varchar (200) | text255 | ADDRESS es un dominio grande, pero tiene muchas palabras repetidas , como Street, Avenue, North, South, etcétera. Las codificac iones text255 y text |

| Columna | Tipo de datos: | Codificación | Explicación |
| --- | --- | --- | --- |
| | | | 32k son útiles para comprimir columnas VARCHAR en las que se repiten las mismas palabras. La longitud de la columna es corta, por lo que text255 es una buena opción. |
| CITY | varchar (30) | text255 | CITY es un dominio grande, con algunos valores repetidos. Algunos nombres de ciudades se utilizan con mucho más frecuencia que otros. Text255 es una buena opción por los mismos motivos que lo es para ADDRESS. |

| Columna | Tipo de datos: | Codificación | Explicación |
|---|---|---|---|
| STATE | char (2) | raw | En los Estados Unidos, STATE es un dominio preciso de 50 valores de dos caracteres. La codificación bytedict generaría cierta compresión, pero como el tamaño de la columna es de solo dos caractere s, es posible que la compresión no compense la sobrecarga de descomprimir los datos. |
| ZIPCODE | char (5) | bytedict | ZIPCODE es un dominio conocido de menos de 50 000 valores únicos. Algunos códigos postales se utilizan con mucho más frecuencia que otros. La codificación bytedict es muy eficaz cuando una columna tiene una cantidad limitada de valores únicos. |

| Columna | Tipo de datos: | Codificación | Explicación |
| --- | --- | --- | --- |
| START_DATE | date | delta32k | Las codificaciones delta son de gran utilidad para las columnas con formato de fecha y hora, en especial si las filas se cargan en el orden de las fechas. |

# Distribución de datos para la optimización de consultas

Cuando carga datos en una tabla, Amazon Redshift distribuye las filas de la tabla a cada uno de los nodos de informática, en función del estilo de distribución de la tabla. Cuando ejecuta una consulta, el optimizador de consultas redistribuye las filas a los nodos de computación según se necesite para realizar combinaciones y agregaciones. El objetivo al elegir un estilo de distribución de tablas es reducir el impacto del paso de redistribución al localizar los datos en el lugar que deben estar antes de que se ejecute la consulta.

> **Note**
>
> En esta sección, se presentarán los principios de la distribución de datos en una base de datos de Amazon Redshift. Se recomienda que cree sus tablas con `DISTSTYLE AUTO`. De hacerlo así, Amazon Redshift utiliza la optimización automática de tablas para elegir el estilo de distribución de los datos. Para obtener más información, consulte Optimización de tablas automática. En el resto de esta sección, se proporcionan detalles sobre los estilos de distribución.

Temas

- Conceptos de distribución de datos
- Estilos de distribución
- Visualización de los estilos de distribución
- Evaluación de los patrones de consulta

- Designación de estilos de distribución
- Evaluación del plan de consulta
- Ejemplo de plan de consulta
- Ejemplos de distribución


## Conceptos de distribución de datos

A continuación, se exponen algunos conceptos de distribución de datos para Amazon Redshift.

### Nodos y sectores

Un clúster de Amazon Redshift es un conjunto de nodos. Cada nodo del clúster tiene su propio sistema operativo, su propia memoria especializada y almacenamiento especializado en el disco. Hay un nodo que es el nodo principal, que administra la distribución de los datos y las tareas de procesamiento de consultas a los nodos de informática. Los nodos de informática proporcionan los recursos necesarios para realizar esas tareas.

El almacenamiento en disco de un nodo de computación se divide en una serie de sectores. El número de sectores por nodo depende del tamaño de nodo del clúster. Todos los nodos participan en la ejecución de consultas en paralelo y trabajan en datos que se distribuyen de la manera más uniforme posible entre los sectores. Para obtener más información acerca de la cantidad de sectores que tiene cada tamaño de nodo, consulte Acerca de clústeres y nodos en la Guía de administración de Amazon Redshift.

### Redistribución de datos

Cuando carga datos en una tabla, Amazon Redshift distribuye las filas de la tabla a cada uno de los sectores del nodo, en función del estilo de distribución de la tabla. Como parte de un plan de consulta, el optimizador determina dónde se deben ubicar los bloques de datos para ejecutar la consulta de la mejor manera. A continuación, los datos se mueven o se redistribuyen físicamente mientras se ejecuta la consulta. La redistribución puede implicar enviar filas específicas a los nodos para realizar una combinación, o bien para difundir una tabla entera a todos los nodos.

La redistribución de datos puede representar una parte sustancial del costo de un plan de consulta y el tráfico de red que genera puede afectar otras operaciones de la base de datos y ralentizar el rendimiento general del sistema. En la medida en que pueda prever dónde conviene ubicar los datos inicialmente, puede reducir el impacto de la redistribución de datos.

### Objetivos de la distribución de datos

Cuando carga datos en una tabla, Amazon Redshift distribuye las filas de la tabla a los nodos de informática y a los sectores, en función del estilo de distribución que se haya elegido al crearla. La distribución de datos tiene dos objetivos principales:

- Distribuir la carga de trabajo de manera uniforme entre los nodos del clúster. Una distribución irregular, o un sesgo en la distribución de datos, exige que algunos nodos trabajen más que otros, lo que perjudica el rendimiento de las consultas.
- Para minimizar el movimiento de datos durante la ejecución de una consulta. Si las filas que participan de uniones o de agregaciones ya están ubicadas junto a los nodos con las filas de las otras tablas con las que se unirán, el optimizador no necesita redistribuir tantos datos cuando se ejecutan las consultas.

La estrategia de distribución que selecciona para su base de datos tiene consecuencias importantes en el rendimiento de la consulta, en los requisitos de almacenamiento, en la carga de datos y en el mantenimiento. Si selecciona el mejor estilo de distribución para cada tabla, puede equilibrar la distribución de sus datos y mejorar considerablemente el rendimiento general del sistema.

## Estilos de distribución

Cuando crea una tabla, puede designar uno de los siguientes estilos de distribución: AUTO, EVEN, KEY o ALL.

Si no se especifica un estilo de distribución, Amazon Redshift usa la distribución AUTO.

Distribución AUTO

Con la distribución AUTO, Amazon Redshift asigna un estilo de distribución óptimo basado en el tamaño de los datos de la tabla. Por ejemplo, si se especifica el estilo de distribución AUTO, Amazon Redshift asigna inicialmente el estilo de distribución ALL a una tabla pequeña. Cuando la tabla crezca, Amazon Redshift podría cambiar el estilo de distribución a KEY y elegir la clave principal (o una columna de la clave primaria compuesta) como la clave de distribución. Si la tabla crece y ninguna de las columnas es adecuada para ser la clave de distribución, Amazon Redshift cambia el estilo de distribución a EVEN. El cambio en el estilo de distribución se produce en segundo plano y tiene un impacto mínimo en las consultas de los usuarios.

Si desea consultar las acciones que Amazon Redshift realizó de forma automática para modificar la clave de distribución de una tabla, consulte SVL_AUTO_WORKER_ACTION. Si desea conocer las recomendaciones actuales relativas a la modificación de la clave de distribución de una tabla, consulte SVV_ALTER_TABLE_RECOMMENDATIONS.

Para ver el estilo de distribución aplicado a una tabla, consulte la vista de catálogo del sistema PG_CLASS_INFO. Para obtener más información, consulte Visualización de los estilos de distribución. Si no se especifica un estilo de distribución con la instrucción CREATE TABLE, Amazon Redshift aplica la distribución AUTO.

Distribución EVEN

El nodo principal distribuye las filas entre los sectores con un método de turnos rotativos, independientemente de los valores de cualquier columna en particular. La distribución EVEN resulta adecuada cuando una tabla no participa en las uniones. Además, resulta adecuada en los casos en que no hay una opción clara entre la distribución KEY y la distribución ALL.

Distribución KEY

Las filas se distribuyen según los valores de una columna. El nodo principal ubica juntos los valores que coinciden en el mismo sector del nodo. Si distribuye un par de tablas en las claves de unión, el nodo principal ubica las filas en los sectores según los valores de las columnas de unión. De este modo, los valores que coinciden de las columnas que tienen en común se almacenan juntos físicamente.

Distribución ALL

Se distribuye una copia de toda la tabla a cada nodo. Mientras que la distribución EVEN o la distribución KEY colocan solo una parte de las filas de la tabla en cada nodo, la distribución ALL garantiza que se coloque cada fila para cada combinación en la que participa la tabla.

La distribución ALL multiplica el almacenamiento requerido por la cantidad de nodos del clúster, por lo que demanda más tiempo para cargar, actualizar o insertar datos en distintas tablas. La distribución ALL es adecuada solo para tablas con movimientos relativamente lentos, es decir tablas que no se actualizan con frecuencia ni de forma generalizada. Dado que el costo de redistribuir tablas pequeñas durante una consulta es bajo, no hay un beneficio significativo para definir tablas de dimensiones pequeñas como DISTSTYLE ALL.

> **Note**
>
> Una vez que haya especificado un estilo de distribución para una columna, Amazon Redshift se encarga de la distribución de datos en el nivel del clúster. Amazon Redshift no requiere ni admite el concepto de partición de datos dentro de los objetos de la base de datos. No es necesario crear espacios de tablas ni definir esquemas de partición para las tablas.

En ciertas situaciones, puede cambiar el estilo de distribución de una tabla después de crearla. Para obtener más información, consulte ALTER TABLE. En aquellas situaciones en las que no puede cambiar el estilo de distribución de una tabla después de crearla, puede volver a crear la tabla y rellenar la nueva tabla con una copia profunda. Para obtener más información, consulte Realización de una copia profunda

## Visualización de los estilos de distribución

Para ver el estilo de distribución de una tabla, consulte la vista PG_CLASS_INFO o la vista SVV_TABLE_INFO.

La columna RELEFFECTIVEDISTSTYLE en PG_CLASS_INFO indica el estilo de distribución actual de la tabla. Si la tabla usa una distribución automática, RELEFFECTIVEDISTSTYLE es 10, 11 o 12, lo que indica que el estilo de distribución efectivo es AUTO (ALL) o AUTO (EVEN) o AUTO (KEY). Si la tabla usa distribución automática, el estilo de distribución podría inicialmente mostrar AUTO (ALL), para a continuación mostrar AUTO (EVEN) o AUTO (KEY) cuando la tabla crezca.

En la siguiente tabla, se proporciona el estilo de distribución para cada valor de la tabla RELEFFECTIVEDISTSTYLE:

| RELEFFECTIVEDISTSTYLE | Estilo de distribución actual |
|---|---|
| 0 | EVEN |
| 1 | KEY |
| 8 | ALL |
| 10 | AUTO (ALL) |
| 11 | AUTO (EVEN) |
| 12 | AUTO (KEY) |

La columna DISTSTYLE en SVV_TABLE_INFO indica el estilo de distribución actual de la tabla. Si la tabla usa una distribución automática, DISTSTYLE es AUTO (ALL), AUTO (EVEN) o AUTO (KEY).

En el siguiente ejemplo, se crean cuatro tablas usando los tres estilos de distribución y distribución automática, luego, se consulta la tabla SVV_TABLE_INFO para ver los estilos de distribución.

```
create table public.dist_key (col1 int)
diststyle key distkey (col1);

insert into public.dist_key values (1);

create table public.dist_even (col1 int)
diststyle even;

insert into public.dist_even values (1);

create table public.dist_all (col1 int)
diststyle all;

insert into public.dist_all values (1);

create table public.dist_auto (col1 int);

insert into public.dist_auto values (1);

select "schema", "table", diststyle from SVV_TABLE_INFO
where "table" like 'dist%';

        schema   |    table        | diststyle
     ------------+-----------------+------------
      public     | dist_key        | KEY(col1)
      public     | dist_even       | EVEN
      public     | dist_all        | ALL
      public     | dist_auto       | AUTO(ALL)
```

## Evaluación de los patrones de consulta

Seleccionar los estilos de distribución es solo un aspecto del diseño de la base de datos. Considere los estilos de distribución en el contexto de todo el sistema, equilibrando la distribución con otros factores importantes como el tamaño del clúster, los métodos de codificación de compresión, las claves de ordenación y las restricciones de tablas.

Pruebe su sistema con datos que sean lo más próximos posibles a los datos reales.

Para poder tomar buenas decisiones respecto a los estilos de distribución, debe comprender los patrones de consulta para su aplicación de Amazon Redshift. Identifique las consultas más costosas de su sistema y base el diseño inicial de su base de datos en las exigencias de esas consultas. Entre

los factores que determinan el costo total de una consulta, se encuentran el tiempo de ejecución de esta y la cantidad de recursos de informática que consume. Asimismo, otros factores que determinan el costo de la consulta son la frecuencia con la que se ejecuta y el grado de interrupción que supone para las demás consultas y operaciones de la base de datos.

Identifique las tablas que utilizan las consultas más costosas y evalúe su función en el tiempo de ejecución de las consultas. Tenga en cuenta cómo se combinan y agregan las tablas.

Utilice las directrices de esta sección para seleccionar un estilo de distribución para cada tabla. Cuando lo haya hecho, cree las tablas y cárguelas con datos que sean lo más próximos posible a los reales. A continuación, pruebe las tablas para los tipos de consultas que prevé utilizar. Puede evaluar los planes de explicación de consultas para identificar oportunidades de ajuste. Compare los tiempos de carga, el espacio de almacenamiento y los tiempos de ejecución de las consultas para poder equilibrar los requisitos generales de su sistema.

## Designación de estilos de distribución

Las consideraciones y recomendaciones de esta sección para designar estilos de distribución utilizan un esquema Star como ejemplo. El diseño de la base de datos se debe basar en un esquema en estrella, en una variante de este esquema o en un esquema completamente diferente. Amazon Redshift está diseñado para funcionar de forma eficaz con cualquier diseño de esquema que elija. Los principios de esta sección se pueden aplicar a cualquier esquema de diseño.

1. Especifique la clave principal y las claves externas de todas las tablas.

   Amazon Redshift no impone restricciones de claves principales y externas, pero el optimizador de consultas las utiliza cuando genera planes de consulta. Si configura claves principales y externas, su aplicación debe conservar la validez de las claves.
2. Distribuya la tabla de hechos y la tabla de mayor dimensión en sus columnas comunes.

   Seleccione la tabla de mayor dimensión en función del tamaño del conjunto de datos que participa en la combinación más frecuente, no solo del tamaño de la tabla. Si hay una tabla que se filtra con frecuencia, con la cláusula WHERE, solo una parte de sus filas participan de la combinación. Dicha tabla tiene menos impacto en la redistribución que una tabla más pequeña que aporta más datos. Designe la clave principal de la tabla de dimensión y la clave externa correspondiente a la tabla de hechos como las claves DISTKEY. Si hay distintas tablas que usan la misma clave de distribución, también se ubican junto a la tabla de hechos. Su tabla de hechos solo puede tener una clave de distribución. Ninguna de las tablas que se unen a otra clave se ubica junto a la tabla de hechos.

3. Designe las claves de distribución para las demás tablas de dimensión.

   Distribuya las tablas en sus claves principales o externas, según cómo se combinan con mayor frecuencia con otras tablas.

4. Valore si conviene cambiar algunas de las tablas de dimensión para utilizar la distribución ALL.

   Si no se puede colocar una tabla de dimensión junto con la tabla de hechos u otra tabla de combinación de importancia, puede mejorar el rendimiento de las consultas de forma significativa distribuyendo la tabla completa a todos los nodos. El uso de la distribución ALL multiplica los requisitos de espacio de almacenamiento y aumenta los tiempos de carga y las operaciones de mantenimiento, por lo que debe analizar todos los factores antes de seleccionar la distribución ALL. En la siguiente sección, se explica cómo identificar posibles candidatos para la distribución ALL mediante una evaluación del plan EXPLAIN.

5. Utilice la distribución AUTO para las tablas restantes.

   Si una tabla se desnormaliza en gran parte y no participa en combinaciones o si no tiene una selección clara respecto a otro estilo de distribución, utilice la distribución AUTO.

Para permitir a Amazon Redshift elegir el estilo de distribución adecuado, no especifique explícitamente un estilo de distribución.

## Evaluación del plan de consulta

Puede utilizar planes de consulta para identificar candidatos para optimizar el estilo de distribución.

Después de tomar las primeras decisiones de diseño, cree sus tablas, cárguelas con datos y pruébelas. Utilice un conjunto de datos de prueba que sean lo más próximos posibles a los datos reales. Mida los tiempos de carga para utilizarlos como referencia para establecer comparaciones.

Evalúe las consultas que son representativas de las consultas más costosas que prevé ejecutar, específicamente aquellas que usan uniones y agregaciones. Compare los tiempos de ejecución de las diferentes opciones de diseño. Cuando compare los tiempos de ejecución, no tenga en cuenta la primera vez que se ejecuta la consulta, ya que el tiempo de la primera ejecución incluye el tiempo de compilación.

DS_DIST_NONE

No es necesario redistribuir porque los sectores correspondientes se ubican junto a los nodos de computación. Por lo general, solo tiene un paso DS_DIST_NONE, la unión entre la tabla de hechos y una tabla de dimensión.

DS_DIST_ALL_NONE

No es necesario redistribuir, porque la tabla de combinación interna utilizó el atributo DISTSTYLE ALL. Se ubica la tabla entera en cada nodo.

DS_DIST_INNER

Se redistribuye la tabla interna.

DS_DIST_OUTER

Se redistribuye la tabla externa.

DS_BCAST_INNER

Se difunde una copia de toda la tabla interna a todos los nodos de computación.

DS_DIST_ALL_INNER

Se redistribuye toda la tabla interna a un único sector porque la tabla externa utiliza el atributo DISTSTYLE ALL.

DS_DIST_BOTH

Ambas tablas se redistribuyen.

DS_DIST_NONE y DS_DIST_ALL_NONE son buenos indicadores. Indican que no fue necesario redistribuir para ese paso porque todas las combinaciones están ubicadas juntas.

DS_DIST_INNER implica que el paso probablemente tiene un costo bastante alto debido a que se redistribuye la tabla interna a los nodos. DS_DIST_INNER indica que la tabla externa ya está distribuida correctamente en la clave de combinación. Configure la clave de distribución de la tabla interna con la clave de combinación para convertirla en DS_DIST_NONE. En algunos casos, no es posible distribuir la tabla interna en la clave de unión dado que la tabla externa no está distribuida en la clave de unión. En este caso, se debe evaluar la posibilidad de utilizar la distribución ALL para la tabla interna. Si la tabla no se actualiza con frecuencia ni de forma generalizada, y tiene el tamaño suficiente para asumir un costo elevado de redistribución, se debe cambiar el estilo de distribución

por ALL y volver a realizar la prueba. La distribución ALL genera un aumento en los tiempos de carga, por lo que cuando vuelva a hacer la prueba debe incluir el tiempo de carga en los factores de evaluación.

DS_DIST_ALL_INNER no es un buen indicador. Significa que la tabla interna completa se redistribuyó a un único sector debido a que la tabla externa usa DISTSTYLE ALL, por lo que se coloca una copia de la tabla externa completa en cada nodo. Esto genera un tiempo de ejecución en serie ineficaz de la combinación en un único nodo, en lugar de sacar provecho del tiempo de ejecución en paralelo utilizando todos los nodos. DISTSTYLE ALL está diseñado para utilizarse solo para la tabla de combinación interna. En lugar de ese atributo, especifique una clave de distribución o utilice una distribución uniforme para la tabla externa.

DS_BCAST_INNER y DS_DIST_BOTH no son bueno indicadores. Por lo general, estas redistribuciones ocurren porque las tablas no están combinadas en sus claves de distribución. Si la tabla de hechos aún no tiene una clave de distribución, especifique la columna de combinación como la clave de distribución para ambas tablas. Si la tabla de hechos ya tiene una clave de distribución en otra columna, debe evaluar si cambiar la clave de distribución para ubicar esta unión mejorará el rendimiento general. Si cambiar la clave de distribución de la tabla externa no es la mejor opción, puede lograr que se ubique mediante la especificación DISTSTYLE ALL para la tabla interna.

En el siguiente ejemplo, se muestra una parte de un plan de consulta con etiquetas DS_BCAST_INNER y DS_DIST_NONE.

```
->  XN Hash Join DS_BCAST_INNER  (cost=112.50..3272334142.59 rows=170771 width=84)
        Hash Cond: ("outer".venueid = "inner".venueid)
        ->  XN Hash Join DS_BCAST_INNER  (cost=109.98..3167290276.71 rows=172456
 width=47)
              Hash Cond: ("outer".eventid = "inner".eventid)
              ->  XN Merge Join DS_DIST_NONE  (cost=0.00..6286.47 rows=172456 width=30)
                    Merge Cond: ("outer".listid = "inner".listid)
                    ->  XN Seq Scan on listing  (cost=0.00..1924.97 rows=192497
 width=14)
                    ->  XN Seq Scan on sales  (cost=0.00..1724.56 rows=172456 width=24)
```

Después de cambiar las tablas de dimensión para usar DISTSTYLE ALL, el plan de consulta para la misma consulta muestra DS_DIST_ALL_NONE en lugar de DS_BCAST_INNER. Además, hay un cambio drástico en el costo relativo de los pasos de combinación. El coste total es `14142.59` en comparación con `3272334142.59` de la consulta anterior.

```
->  XN Hash Join DS_DIST_ALL_NONE  (cost=112.50..14142.59 rows=170771 width=84)
```

```
        Hash Cond: ("outer".venueid = "inner".venueid)
        ->  XN Hash Join DS_DIST_ALL_NONE  (cost=109.98..10276.71 rows=172456 width=47)
              Hash Cond: ("outer".eventid = "inner".eventid)
              ->  XN Merge Join DS_DIST_NONE  (cost=0.00..6286.47 rows=172456 width=30)
                    Merge Cond: ("outer".listid = "inner".listid)
                    ->  XN Seq Scan on listing  (cost=0.00..1924.97 rows=192497
 width=14)
                    ->  XN Seq Scan on sales  (cost=0.00..1724.56 rows=172456 width=24)
```

## Ejemplo de plan de consulta

En este ejemplo, se muestra cómo evaluar un plan de consulta para encontrar oportunidades y optimización de las distribuciones.

Ejecute la siguiente consulta con un comando EXPLAIN para producir un plan de consulta.

```
explain
select lastname, catname, venuename, venuecity, venuestate, eventname,
month, sum(pricepaid) as buyercost, max(totalprice) as maxtotalprice
from category join event on category.catid = event.catid
join venue on venue.venueid = event.venueid
join sales on sales.eventid = event.eventid
join listing on sales.listid = listing.listid
join date on sales.dateid = date.dateid
join users on users.userid = sales.buyerid
group by lastname, catname, venuename, venuecity, venuestate, eventname, month
having sum(pricepaid)>9999
order by catname, buyercost desc;
```

En la base de datos TICKIT, SALES es una tabla de hechos y LISTING la tabla de mayor dimensión. Para poder ubicar juntas las tablas, SALES se distribuye en LISTID, que es la clave externa para LISTING, y LISTING se distribuye en su clave principal, LISTID. En el siguiente ejemplo, se muestran los comandos CREATE TABLE para SALES y LISTING.

```
create table sales(
 salesid integer not null,
 listid integer not null distkey,
 sellerid integer not null,
 buyerid integer not null,
 eventid integer not null encode mostly16,
 dateid smallint not null,
```

```
 qtysold smallint not null encode mostly8,
 pricepaid decimal(8,2) encode delta32k,
 commission decimal(8,2) encode delta32k,
 saletime timestamp,
 primary key(salesid),
 foreign key(listid) references listing(listid),
 foreign key(sellerid) references users(userid),
 foreign key(buyerid) references users(userid),
 foreign key(dateid) references date(dateid))
       sortkey(listid,sellerid);

create table listing(
 listid integer not null distkey sortkey,
 sellerid integer not null,
 eventid integer not null encode mostly16,
 dateid smallint not null,
 numtickets smallint not null encode mostly8,
 priceperticket decimal(8,2) encode bytedict,
 totalprice decimal(8,2) encode mostly32,
 listtime timestamp,
 primary key(listid),
 foreign key(sellerid) references users(userid),
 foreign key(eventid) references event(eventid),
 foreign key(dateid) references date(dateid));
```

En el siguiente plan de consulta, el paso de combinación de fusión para la combinación de SALES y LISTING muestra la etiqueta DS_DIST_NONE, que indica que no es necesario redistribuir en este paso. No obstante, si avanzamos en el plan de consulta, las otras combinaciones internas muestran la etiqueta DS_BCAST_INNER, que indica que la tabla interna se difunde como parte de la ejecución de la consulta. Como con la distribución de clave solo se pueden ubicar juntas un par de tablas, se deben volver a difundir cinco tablas.

```
QUERY PLAN
XN Merge  (cost=1015345167117.54..1015345167544.46 rows=1000 width=103)
  Merge Key: category.catname, sum(sales.pricepaid)
  ->  XN Network  (cost=1015345167117.54..1015345167544.46 rows=170771 width=103)
        Send to leader
        ->  XN Sort  (cost=1015345167117.54..1015345167544.46 rows=170771 width=103)
              Sort Key: category.catname, sum(sales.pricepaid)
              ->  XN HashAggregate  (cost=15345150568.37..15345152276.08 rows=170771
 width=103)
                    Filter: (sum(pricepaid) > 9999.00)
```

```
                     ->  XN Hash Join DS_BCAST_INNER  (cost=742.08..15345146299.10
rows=170771 width=103)
                           Hash Cond: ("outer".catid = "inner".catid)
                           ->  XN Hash Join DS_BCAST_INNER
(cost=741.94..15342942456.61 rows=170771 width=97)
                                 Hash Cond: ("outer".dateid = "inner".dateid)
                                 ->  XN Hash Join DS_BCAST_INNER
(cost=737.38..15269938609.81 rows=170766 width=90)
                                       Hash Cond: ("outer".buyerid = "inner".userid)
                                       ->  XN Hash Join DS_BCAST_INNER
(cost=112.50..3272334142.59 rows=170771 width=84)
                                             Hash Cond: ("outer".venueid =
"inner".venueid)
                                             ->  XN Hash Join DS_BCAST_INNER
(cost=109.98..3167290276.71 rows=172456 width=47)
                                                   Hash Cond: ("outer".eventid =
"inner".eventid)
                                                   ->  XN Merge Join DS_DIST_NONE
(cost=0.00..6286.47 rows=172456 width=30)
                                                         Merge Cond: ("outer".listid =
"inner".listid)
                                                         ->  XN Seq Scan on listing
(cost=0.00..1924.97 rows=192497 width=14)
                                                         ->  XN Seq Scan on sales
(cost=0.00..1724.56 rows=172456 width=24)
                                                   ->  XN Hash  (cost=87.98..87.98
rows=8798 width=25)
                                                         ->  XN Seq Scan on event
(cost=0.00..87.98 rows=8798 width=25)
                                             ->  XN Hash  (cost=2.02..2.02 rows=202
width=41)
                                                   ->  XN Seq Scan on venue
(cost=0.00..2.02 rows=202 width=41)
                                       ->  XN Hash  (cost=499.90..499.90 rows=49990
width=14)
                                             ->  XN Seq Scan on users
(cost=0.00..499.90 rows=49990 width=14)
                                 ->  XN Hash  (cost=3.65..3.65 rows=365 width=11)
                                       ->  XN Seq Scan on date  (cost=0.00..3.65
rows=365 width=11)
                           ->  XN Hash  (cost=0.11..0.11 rows=11 width=10)
                                 ->  XN Seq Scan on category  (cost=0.00..0.11 rows=11
width=10)
```

Una solución es modificar las tablas para que tengan DISTSTYLE ALL.

```
ALTER TABLE users ALTER DISTSTYLE ALL;
ALTER TABLE venue ALTER DISTSTYLE ALL;
ALTER TABLE category ALTER DISTSTYLE ALL;
ALTER TABLE date ALTER DISTSTYLE ALL;
ALTER TABLE event ALTER DISTSTYLE ALL;
```

Ejecute nuevamente la misma consulta con el comando EXPLAIN y examine el plan de consulta nuevo. Las combinaciones ahora muestran la etiqueta DS_DIST_ALL_NONE, lo que indica que no es necesario redistribuir porque ya se distribuyeron los datos a cada nodo con el atributo DISTSTYLE ALL.

```
QUERY PLAN
XN Merge  (cost=1000000047117.54..1000000047544.46 rows=1000 width=103)
  Merge Key: category.catname, sum(sales.pricepaid)
  ->  XN Network  (cost=1000000047117.54..1000000047544.46 rows=170771 width=103)
        Send to leader
        ->  XN Sort  (cost=1000000047117.54..1000000047544.46 rows=170771 width=103)
              Sort Key: category.catname, sum(sales.pricepaid)
              ->  XN HashAggregate  (cost=30568.37..32276.08 rows=170771 width=103)
                    Filter: (sum(pricepaid) > 9999.00)
                    ->  XN Hash Join DS_DIST_ALL_NONE  (cost=742.08..26299.10
 rows=170771 width=103)
                          Hash Cond: ("outer".buyerid = "inner".userid)
                          ->  XN Hash Join DS_DIST_ALL_NONE  (cost=117.20..21831.99
 rows=170766 width=97)
                                Hash Cond: ("outer".dateid = "inner".dateid)
                                ->  XN Hash Join DS_DIST_ALL_NONE
 (cost=112.64..17985.08 rows=170771 width=90)
                                      Hash Cond: ("outer".catid = "inner".catid)
                                      ->  XN Hash Join DS_DIST_ALL_NONE
 (cost=112.50..14142.59 rows=170771 width=84)
                                            Hash Cond: ("outer".venueid =
 "inner".venueid)
                                            ->  XN Hash Join DS_DIST_ALL_NONE
 (cost=109.98..10276.71 rows=172456 width=47)
                                                  Hash Cond: ("outer".eventid =
 "inner".eventid)
                                                  ->  XN Merge Join DS_DIST_NONE
 (cost=0.00..6286.47 rows=172456 width=30)
                                                        Merge Cond: ("outer".listid =
 "inner".listid)
```

```
                                                        ->  XN Seq Scan on listing
 (cost=0.00..1924.97 rows=192497 width=14)
                                                        ->  XN Seq Scan on sales
 (cost=0.00..1724.56 rows=172456 width=24)
                                                  ->  XN Hash  (cost=87.98..87.98
 rows=8798 width=25)
                                                        ->  XN Seq Scan on event
 (cost=0.00..87.98 rows=8798 width=25)
                                              ->  XN Hash  (cost=2.02..2.02 rows=202
 width=41)
                                                    ->  XN Seq Scan on venue
 (cost=0.00..2.02 rows=202 width=41)
                                        ->  XN Hash  (cost=0.11..0.11 rows=11 width=10)
                                              ->  XN Seq Scan on category
 (cost=0.00..0.11 rows=11 width=10)
                                  ->  XN Hash  (cost=3.65..3.65 rows=365 width=11)
                                        ->  XN Seq Scan on date  (cost=0.00..3.65
 rows=365 width=11)
                            ->  XN Hash  (cost=499.90..499.90 rows=49990 width=14)
                                  ->  XN Seq Scan on users  (cost=0.00..499.90 rows=49990
 width=14)
```

## Ejemplos de distribución

En los siguientes ejemplos, se muestra cómo se distribuyeron los datos según las opciones que definió en la instrucción CREATE TABLE.

### Ejemplos de DISTKEY

Observe el esquema de la tabla USERS en la base de datos TICKIT. USERID se define como la columna SORTKEY y la columna DISTKEY:

```
select "column", type, encoding, distkey, sortkey
from pg_table_def where tablename = 'users';

    column     |          type          | encoding | distkey | sortkey
---------------+------------------------+----------+---------+---------
 userid        | integer                | none     | t       |       1
 username      | character(8)           | none     | f       |       0
 firstname     | character varying(30)  | text32k  | f       |       0

...
```

USERID es una buena opción para la columna de distribución de esta tabla. Si consulta la vista de sistema SVV_DISKUSAGE, puede ver que la tabla está bien distribuida de manera uniforme. Los números de las columnas comienzan con cero, por lo que USERID es la columna 0.

```
select slice, col, num_values as rows, minvalue, maxvalue
from svv_diskusage
where name='users' and col=0 and rows>0
order by slice, col;

slice| col | rows  | minvalue | maxvalue
-----+-----+-------+----------+----------
0    | 0   | 12496 | 4        | 49987
1    | 0   | 12498 | 1        | 49988
2    | 0   | 12497 | 2        | 49989
3    | 0   | 12499 | 3        | 49990
(4 rows)
```

La tabla tiene 49 990 filas. En la columna rows (num_values), se ve que cada sector tiene, aproximadamente, la misma cantidad de filas. En las columnas minvalue y maxvalue, se ve el rango de valores en cada sector. Cada sector incluye prácticamente todo el rango de valores, por lo que es muy probable que todos los sectores participen de la ejecución de una consulta que filtre en busca de un rango de ID de usuarios.

En este ejemplo, se muestra la distribución en un sistema pequeño de prueba. Por lo general, la cantidad total de sectores es mucho mayor.

Si suele hacer combinaciones o agrupaciones con la columna STATE, puede seleccionar distribuir en la columna STATE. En el siguiente ejemplo, se muestra un caso en el que se crea una tabla nueva con los mismos datos que la tabla USERS, pero se establece DISTKEY en la columna STATE. En este caso, la distribución no es tan uniforme. El sector 0 (13 587 filas) dispone, aproximadamente, de un 30 % más de filas que el sector 3 (10 150 filas). En una tabla mucho más grande, este nivel de sesgo de distribución puede tener un impacto negativo en el procesamiento de las consultas.

```
create table userskey distkey(state) as select * from users;

select slice, col, num_values as rows, minvalue, maxvalue from svv_diskusage
where name = 'userskey' and col=0 and rows>0
order by slice, col;

slice | col | rows  | minvalue | maxvalue
------+-----+-------+----------+----------
```

```
    0 |   0 | 13587 |        5 |    49989
    1 |   0 | 11245 |        2 |    49990
    2 |   0 | 15008 |        1 |    49976
    3 |   0 | 10150 |        4 |    49986
(4 rows)
```

## Ejemplo de DISTSTYLE EVEN

Si crea una nueva tabla con los mismos datos que la tabla USERS, pero configura DISTSTYLE como EVEN, las filas siempre se distribuyen de manera uniforme en todos los sectores.

```
create table userseven diststyle even as
select * from users;

select slice, col, num_values as rows, minvalue, maxvalue from svv_diskusage
where name = 'userseven' and col=0 and rows>0
order by slice, col;

slice | col | rows  | minvalue | maxvalue
------+-----+-------+----------+----------
    0 |   0 | 12497 |        4 |    49990
    1 |   0 | 12498 |        8 |    49984
    2 |   0 | 12498 |        2 |    49988
    3 |   0 | 12497 |        1 |    49989
(4 rows)
```

No obstante, como la distribución no está basada en una columna específica, el procesamiento de consultas puede empeorar, en especial si la tabla se combina con otras tablas. La falta de distribución en una columna de combinación suele influir en el rendimiento del tipo de operación de combinación por ejecutar. Las operaciones de combinación, agregación y agrupamiento están optimizadas cuando se distribuyen y ordenan ambas tablas en sus respectivas columnas de combinación.

## Ejemplo de DISTSTYLE ALL

Si crea una tabla nueva con los mismos datos que la tabla USERS, pero configura DISTSTYLE como ALL, todas las filas se distribuyen al primer sector de cada nodo.

```
select slice, col, num_values as rows, minvalue, maxvalue from svv_diskusage
where name = 'usersall' and col=0 and rows > 0
order by slice, col;
```

```
slice | col | rows  | minvalue | maxvalue
------+-----+-------+----------+----------
    0 |   0 | 49990 |        4 |    49990
    2 |   0 | 49990 |        2 |    49990

(4 rows)
```

# Claves de clasificación

> **Note**
>
> Se recomienda que cree sus tablas con `SORTKEY AUTO`. En este caso, Amazon Redshift utiliza la optimización automática de tablas para elegir la clave de ordenación. Para obtener más información, consulte Optimización de tablas automática. En el resto de esta sección, se proporcionan detalles sobre el orden.

Cuando se crea una tabla, se pueden definir, como alternativa, una o varias de sus columnas como claves de ordenación. Cuando carga por primera vez los datos en una tabla vacía, las filas se almacenan en el disco de forma ordenada. Se transmite la información relacionada con las columnas con clave de ordenación al planificador de consultas, que utiliza esta información para construir planes que aprovechen la forma en la que se almacenan los datos. Para obtener más información, consulte CREATE TABLE. Para obtener información sobre las prácticas recomendadas a la hora de crear una clave de clasificación, consulte Elección de la clave de clasificación recomendada.

La ordenación permite encargarse eficazmente de predicados restringidos por rangos. Amazon Redshift .almacena los datos de columna en bloques de disco de 1 MB. Los valores mínimo y máximo de cada bloque se almacenan como parte de los metadatos. Si una consulta usa un predicado de rango restringido, el procesador de consultas puede usar los valores mínimos y máximos para omitir rápidamente grandes cantidades de bloques durante los análisis de las tablas. Por ejemplo, supongamos que una tabla almacena cinco años de datos ordenados por fecha y una consulta especifica un rango de fechas correspondiente a un mes. En este caso, se puede eliminar hasta el 98 % de los bloques de disco del análisis. Si los datos no están ordenados, se deben examinar más bloques del disco (posiblemente todos).

Puede especificar una clave de ordenación compuesta o intercalada. Una clave de ordenación compuesta es más eficaz cuando los predicados de la consulta usan un prefijo, que es un subconjunto ordenado de columnas de clave de ordenación. Una clave de ordenación intercalada

le otorga el mismo peso a cada columna de la clave de ordenación, por lo que los predicados de la consulta pueden usar cualquier subconjunto de columnas que conforman la clave de ordenación, en cualquier orden.

Para comprender el impacto de la clave de ordenación seleccionada en el rendimiento de las consultas, utilice el comando EXPLAIN. Para obtener más información, consulte Flujo de trabajo de planificación y ejecución de consultas.

Para definir un tipo de ordenación, utilice la palabra clave INTERLEAVED o COMPOUND en su instrucción CREATE TABLE o CREATE TABLE AS. La opción predeterminada es COMPOUND. Se recomienda el COMPOUND cuando actualice las tablas a menudo con operaciones INSERT, UPDATE o DELETE. Una clave de ordenación INTERLEAVED puede usar como máximo ocho columnas. En función de los datos y del tamaño del clúster, VACUUM REINDEX necesita mucho más tiempo que VACUUM FULL ya que realiza una ejecución adicional para analizar las claves de ordenación intercaladas. La operación de clasificación y fusión tarda más tiempo para tablas intercaladas porque es posible que la clasificación intercalada tenga que reorganizar más filas que una clasificación compuesta.

Para ver las claves de ordenación de una tabla, consulte la vista de sistema SVV_TABLE_INFO.

Temas

- Clasificación del diseño de datos multidimensional (versión preliminar)
- Clave de ordenación compuesta
- Clave de ordenación intercalada

## Clasificación del diseño de datos multidimensional (versión preliminar)

La siguiente es una versión preliminar de la documentación sobre la clasificación del diseño de datos multidimensional de tablas. Tanto la documentación como la característica quedan sujetas a cambios. Se recomienda utilizar esta característica solo con clústeres de prueba y no en entornos de producción. Para obtener una vista previa de los términos y las condiciones, consulte Participación de la versión beta del servicio en Términos del servicio de AWS.

> **Note**
>
> Esta característica solo está disponible en la versión preliminar de los clústeres. Para crear un clúster de versión preliminar, consulte Crear un clúster de versión preliminar en la Guía de administración de Amazon Redshift.

Una clave de clasificación de diseño de datos multidimensional es un tipo de clave de clasificación AUTO que se basa en predicados repetitivos encontrados en una carga de trabajo. Si su carga de trabajo tiene predicados repetitivos, Amazon Redshift puede mejorar el rendimiento del escaneo de tablas mediante la colocación de filas de datos que satisfagan los predicados repetitivos. En lugar de almacenar los datos de una tabla siguiendo un estricto orden de columnas, una clave de clasificación de diseño de datos multidimensional almacena los datos mediante el análisis de los predicados repetitivos que aparecen en una carga de trabajo. Se puede encontrar más de un predicado repetitivo en una carga de trabajo. Según la carga de trabajo, este tipo de clave de clasificación puede mejorar el rendimiento de muchos predicados. Amazon Redshift determina automáticamente si este método de clave de clasificación debe utilizarse para las tablas definidas con una clave de clasificación `AUTO`.

Por ejemplo, supongamos que tiene una tabla con los datos clasificados en orden de columna. Es posible que sea necesario examinar muchos bloques de datos para determinar si satisfacen los predicados de la carga de trabajo. Sin embargo, si los datos se almacenan en el disco siguiendo un orden de predicados, será necesario escanear menos bloques para satisfacer la consulta. En este caso, usar una clave de clasificación de diseño de datos multidimensional es beneficioso.

Para ver si una consulta utiliza una clave de diseño de datos multidimensional, consulte la columna `step_attribute` de la vista SYS_QUERY_DETAIL. Si el valor es `multi-dimensional`, entonces se habrá utilizado un diseño de datos multidimensional para la consulta. Para ver si una tabla definida con la clave de clasificación AUTO usa un diseño de datos multidimensional, consulte la columna `sortkey1` de la vista SVV_TABLE_INFO. Si el valor es `padb_internal_mddl_key_col`, entonces se habrá utilizado un diseño de datos multidimensional para la clave de clasificación de la tabla.

Para evitar que Amazon Redshift utilice una clave de clasificación de diseño de datos multidimensional, elija una opción de clave de clasificación de tablas distinta de `SORTKEY AUTO`. Para obtener más información acerca de las opciones SORTKEY, consulte CREATE TABLE.

## Clave de ordenación compuesta

Una clave compuesta está formada por todas las columnas presentes en la definición de la clave de ordenación, en el orden que aparecen. Una clave de ordenación compuesta es más útil cuando el filtro de una consulta aplica condiciones, como filtros y combinaciones, que usan un prefijo de clave de ordenación. Los beneficios del rendimiento de la ordenación compuesta se reducen cuando las consultas dependen solo de las columnas de ordenación secundarias, sin hacer referencia a las columnas principales. COMPOUND es el tipo de ordenación predeterminado.

Las claves de ordenación compuesta pueden acelerar las combinaciones, las operaciones GROUP BY y ORDER BY y las funciones de ventana que usan PARTITION BY y ORDER BY. Por ejemplo, una combinación de fusión, que suele ser más rápida que una combinación hash, es factible cuando los datos se distribuyen y se ordenan previamente en las columnas de combinación. Las claves de ordenación compuesta también ayudan a mejorar la compresión.

A medida que agrega filas a una tabla ordenada que ya tiene datos, la región desordenada aumenta y esto tiene un efecto significativo en el rendimiento. El efecto es mayor cuando la tabla usa una ordenación intercalada, en especial cuando las columnas de ordenación incluyen datos que aumentan de forma monótona, como las columnas de fecha o de marca temporal. Ejecute una operación VACUUM de manera periódica, sobre todo después de grandes cargas de datos, para volver a clasificar y a analizar los datos. Para obtener más información, consulte Reducción del tamaño de la región no ordenada. Después hacer la limpieza para volver a ordenar los datos, se recomienda ejecutar un comando ANALYZE para actualizar los metadatos estadísticos para el planificados de consultas. Para obtener más información, consulte Análisis de tablas.

## Clave de ordenación intercalada

Una clave de ordenación intercalada le otorga el mismo peso a cada columna o subconjunto de columnas en la clave de ordenación. Si hay distintas consultas que usan diferentes columnas para filtros, puede, por lo general, mejorar el rendimiento de esas consultas utilizando un estilo de ordenación intercalada. Cuando una consulta usa predicados restrictivos en las columnas de ordenación secundarias, la ordenación intercalada mejora considerablemente el rendimiento de la consulta, si se la compara con la ordenación compuesta.

> **Important**
>
> No utilice una clave de ordenación intercalada con atributos monótonamente crecientes, como columnas de identidad, fechas o marcas temporales.

Las mejoras de rendimiento que obtiene al implementar una clave de ordenación intercalada se deben compensar con el aumento de los tiempos de carga y de limpieza.

Las ordenaciones intercaladas son más eficaces con las consultas sumamente selectivas que filtran una o más columnas con clave de ordenación en la cláusula WHERE; por ejemplo, la consulta `select c_name from customer where c_region = 'ASIA'`. Los beneficios de la ordenación intercalada aumentan con la cantidad de columnas ordenadas que están limitadas.

La ordenación intercalada es más eficaz con las tablas grandes. La ordenación se aplica a cada sector. Por lo tanto, una ordenación intercalada resulta más eficaz si una tabla es lo suficientemente grande como para requerir varios bloques de 1 MB por sector. En este caso, el procesador de consultas puede omitir una proporción significativa de los bloques mediante el uso de predicados restrictivos. Para ver la cantidad de bloques que utiliza una tabla, consulte la vista de sistema STV_BLOCKLIST.

Cuando se ordena una sola columna, una ordenación intercalada puede tener un mejor rendimiento que una ordenación compuesta si los valores de la columna tienen un prefijo largo en común. Por ejemplo, los URL, por lo general, comienzan con "http://www". Las claves de ordenación compuesta usan una cantidad limitada de caracteres del prefijo, lo que genera muchas duplicaciones de claves. Las ordenaciones intercaladas utilizan un esquema de compresión interno para los valores de mapas de zonas que les permite discriminar mejor entre valores de columnas que tienen un prefijo largo en común.

Al migrar clústeres aprovisionados de Amazon Redshift a Amazon Redshift sin servidor, Redshift convierte tablas con claves de clasificación intercaladas y DISTSTYLE KEY en claves de clasificación compuestas. El DISTSTYLE no cambia. Para obtener más información sobre los estilos de distribución, consulte Uso de estilos de distribución de datos.

VACUUM REINDEX

A medida que agrega filas a una tabla ordenada que ya tiene datos, el rendimiento puede deteriorarse con el tiempo. Este deterioro ocurre tanto en las ordenaciones compuestas como en las intercaladas, pero tiene mayor efecto en las tablas intercaladas. El comando VACUUM restaura el orden, pero la operación puede tomar más tiempo para las tablas intercaladas porque fusionar datos nuevos intercalados puede implicar modificar cada bloque de datos.

Cuando las tablas se cargan por primera vez, Amazon Redshift analiza la distribución de los valores de las columnas con clave de ordenación y usa la información para intercalar de manera óptima las columnas con clave de ordenación. A medida que una tabla aumenta, la distribución de los

valores de las columnas con clave de ordenación puede variar o generar un sesgo, en especial con las columnas de fecha o de marca temporal. Si el sesgo es demasiado grande, puede afectar el rendimiento. Para volver a analizar las claves de ordenación y restaurar el rendimiento, ejecute el comando VACUUM con la palabra clave REINDEX. Como el comando VACUUM REINDEX debe hacer un análisis adicional sobre los datos, puede tardar más que un comando VACUUM estándar para las tablas intercaladas. Para ver información acerca del sesgo de distribución de claves y del tiempo de la última reindexación, consulte la vista de sistema SVV_INTERLEAVED_COLUMNS.

Para obtener más información acerca de cómo determinar con qué frecuencia ejecutar el comando VACUUM y cuándo ejecutar el comando VACUUM REINDEX, consulte Decisión de si se debe reindexar.

## Restricciones de tablas

Las restricciones de unicidad, clave principal y clave externa son solo informativas, Amazon Redshift no las aplica al rellenar una tabla. Por ejemplo, si inserta datos en una tabla con dependencias, la inserción puede realizarse correctamente aunque infrinja la restricción. No obstante, las claves principales y externas se utilizan como sugerencias de planificación y deben estar declaradas si el proceso de Extract, Transform, Load (ETL, Extracción, transformación y carga) o algún otro proceso de su aplicación exige su integridad.

Por ejemplo, el planificador de consultas emplea claves principales y externas en determinados cálculos estadísticos. Con ello se infiere la unicidad y las relaciones referenciales que afectan a las técnicas de descorrelación de las subconsultas. De este modo, puede ordenar una gran cantidad de uniones y eliminar aquellas que sean redundantes.

El planificador aprovecha estas relaciones entre claves, pero supone que todas las claves de las tablas de Amazon Redshift son válidas tal y como se cargan. Si su aplicación admite claves externa o principales no válidas, algunas consultas podrían devolver resultados incorrectos. Por ejemplo, una consulta SELECT DISTINCT podría devolver filas duplicadas si la clave principal no es única. No defina restricciones de claves para sus tablas si duda de su validez. Sin embargo, siempre debe declarar las claves principales y externas y las restricciones de unicidad si sabe que son válidas.

Amazon Redshift sí impone restricciones en las columnas NOT NULL.

Para obtener más información acerca de estas restricciones, consulte CREATE TABLE. Para obtener información acerca de cómo eliminar una tabla con dependencias, consulte DROP TABLE.

# Carga de datos en Amazon Redshift

Existen varias formas de cargar datos en una base de datos de Amazon Redshift. Un origen popular de datos para cargar son los archivos de Amazon S3. En la siguiente tabla se resumen algunos de los métodos que se utilizarán al empezar desde un origen de Amazon S3.

| Método a utilizar | Descripción | Cuando se necesita un método |
|---|---|---|
| COPY command | Ejecuta una ingesta de archivos por lotes para cargar datos de los archivos de Amazon S3. Este método aprovecha las capacidades de procesamiento en paralelo de Amazon Redshift. Para obtener más información, consulte Carga de tablas con el comando COPY. | Debe usarse cuando se necesiten requisitos básicos de carga de datos para iniciar la ingesta manual de archivos por lotes. Este método se utiliza principalmente con canalizaciones de ingesta de archivos personalizadas y de terceros o cargas de trabajo de ingesta de archivos puntuales. |
| COPY... Comando CREATE JOB (copia automática) | Ejecuta los comandos COPY automáticamente cuando se crea un nuevo archivo en las rutas rastreadas de Amazon S3. Para obtener más información, consulte Creación de una integración de eventos de S3 para copiar automáticamente archivos desde buckets de Amazon S3. | Se debe usar cuando una canalización de ingesta de archivos necesite ingerir datos automáticamente al crear un nuevo archivo en Amazon S3. Amazon Redshift hace un seguimiento de los archivos que se pueden ingerir para evitar la duplicación de datos. Este método requiere que los propietarios de los buckets de Amazon S3 lo configuren. |
| Carga desde consultas de lagos de datos | Cree tablas externas para ejecutar consultas de lagos de datos en los archivos de | Se debe utilizar en cualquiera de los siguientes escenarios: |

| Método a utilizar | Descripción | Cuando se necesita un método |
| --- | --- | --- |
| | Amazon S3 y, a continuación, ejecute el comando INSERT INTO para cargar los resultados de las consultas de los lagos de datos en las tablas locales. Para obtener más información, consulte Tablas externas para Redshift Spectrum. | • Carga de AWS Glue y formatos de tabla abiertos (como Apache Iceberg, Apache Hudi o Delta Lake).<br>• Los archivos de origen se deben ingerir parcialmente (por ejemplo, se necesitan para ejecutar una cláusula WHERE para ingerir determinadas filas).<br>• Se necesita más flexibilidad para ingerir determinadas columnas (como ejecutar un comando SELECT) o realizar transformaciones de datos básicas sobre la marcha (como aplicar operaciones básicas o llamar a las UDF en los valores del archivo de origen). |
| Otros métodos que puede considerar | | |

| Método a utilizar | Descripción | Cuando se necesita un método |
| --- | --- | --- |
| Ingesta de streaming | La ingesta de streaming proporciona una ingesta de alta velocidad y baja latencia de datos de flujos de Amazon Kinesis Data Streams y Amazon Managed Streaming para Apache Kafka en una vista materializada de Amazon Redshift aprovisionado o Redshift sin servidor. Para obtener más información, consulte Introducción a la ingesta de streaming de Amazon Kinesis Data Streams y Introducción a la ingesta de transmisiones desde orígenes de Apache Kafka. | Debe tenerse en cuenta para los casos de uso en los que los datos se transmiten primero a archivos en Amazon S3 y, después, se cargan desde Amazon S3. Si no es necesario conservar los datos en Amazon S3, a menudo puede considerar la posibilidad de transmitirlos directamente a Amazon Redshift. |
| Ejecución de consultas del lago de datos | Ejecución de consultas directamente desde una tabla de un lago de datos en lugar de ingerir el contenido de la tabla a una tabla local. Para obtener más información, consulte Amazon Redshift Spectrum. | Se debe usar cuando el caso de uso no requiera el rendimiento de consultas de tablas locales en Amazon Redshift. |

| Método a utilizar | Descripción | Cuando se necesita un método |
| --- | --- | --- |
| Carga por lotes con el editor de consultas de Amazon Redshift v2 | Puede preparar y ejecutar las cargas de trabajo de ingesta de archivos por lotes visualmente en el editor de consultas de Amazon Redshift v2. Para obtener más información, consulte Carga de datos desde S3 en la Guía de administración de Amazon Redshift. | Se debe usar cuando desee que el editor de consultas v2 prepare las instrucciones COPY y desee una herramienta visual que simplifique el proceso de preparación de las instrucciones COPY. |
| Carga de datos desde un archivo local con el editor de consultas de Amazon Redshift v2 | Puede cargar archivos directamente desde el escritorio a las tablas de Amazon Redshift sin necesidad de cargar los archivos manualmente en Amazon S3. Para obtener más información, consulte Carga de datos desde una configuración de archivo local y flujo de trabajo en la Guía de administración de Amazon Redshift. | Se debe usar cuando necesite cargar archivos rápidamente desde el equipo local para realizar consultas únicas. Con este método, el editor de consultas de Amazon Redshift v2 almacena temporalmente el archivo en un bucket de Amazon S3 propiedad del cliente y ejecuta un comando de copia mediante esta ruta de Amazon S3. |

La forma más eficiente de cargar una tabla es mediante el comando COPY. También puede agregar datos a las tablas mediante el uso de los comandos INSERT, aunque ese método es menos efectivo que usar COPY. El comando COPY puede leer datos de diferentes archivos y flujos de datos de forma simultánea. Amazon Redshift asigna la carga de trabajo a los nodos de Amazon Redshift y realiza las operaciones de carga en paralelo, incluida la ordenación de las filas y la distribución de los datos entre los sectores de los nodos.

> **Note**
>
> Las tablas externas de Amazon Redshift Spectrum solo se pueden leer. No puede utilizar COPY ni INSERT en una tabla externa.

Para acceder a los datos en otros recursos de AWS, Amazon Redshift debe tener permiso para acceder a tales recursos y para realizar las acciones necesarias para acceder a los datos. Puede usar AWS Identity and Access Management (IAM) para limitar el acceso de los usuarios a los datos y recursos de Amazon Redshift.

Tras la carga de datos inicial, si agrega, modifica o elimina una cantidad de datos significativa, a continuación debe ejecutar un comando VACUUM para reorganizar los datos y recuperar espacio tras las eliminaciones. También debe ejecutar un comando ANALYZE para actualizar las estadísticas de la tabla.

Temas


- Carga de tablas con el comando COPY
- Creación de una integración de eventos de S3 para copiar automáticamente archivos desde buckets de Amazon S3
- Carga de tablas con comandos DML
- Realización de una copia profunda
- Análisis de tablas
- Limpieza de tablas
- Administración de operaciones de escritura simultáneas
- Tutorial: Carga de datos desde Amazon S3


# Carga de tablas con el comando COPY

El comando COPY utiliza la arquitectura de procesamiento masivo en paralelo (MPP) de Amazon Redshift para leer y cargar datos en paralelo desde archivos de Amazon S3, desde una tabla de DynamoDB o desde un texto de salida de uno o más alojamientos remotos.

Antes de obtener información sobre todas las opciones del comando COPY, le recomendamos que obtenga información sobre las opciones básicas para cargar datos de Amazon S3. La Guía de introducción de Amazon Redshift muestra un uso sencillo del comando COPY para cargar datos

de Amazon S3 mediante un rol de IAM predeterminado. Consulte Paso 4: Carga de datos desde Amazon S3 a Amazon Redshift para obtener más información.

> **Note**
>
> Le recomendamos encarecidamente que use el comando COPY para cargar grandes cantidades de datos. El uso de instrucciones INSERT individuales para completar una tabla podría ser prohibitivamente lento. O bien, si los datos ya existen en otras tablas de la base de datos de Amazon Redshift, puede utilizar INSERT INTO… SELECT o CREATE TABLE AS para mejorar el rendimiento. Para obtener información, consulte INSERT o CREATE TABLE AS.

Para cargar datos desde otro recurso de AWS, Amazon Redshift debe tener permiso para acceder al recurso y realizar las acciones necesarias.

Para otorgar o revocar el privilegio de carga de datos en una tabla mediante una instrucción COPY, otorgue o revoque el privilegio INSERT.

Los datos deben tener el formato correcto para cargase en la tabla de Amazon Redshift. En esta sección, se presentan las directrices para preparar y controlar sus datos antes de la carga y para validar una instrucción COPY antes de ejecutarla.

Para proteger la información en los archivos, puede cifrar los archivos de datos antes de cargarlos al bucket de Amazon S3; COPY descifrará los datos a medida que realice la carga. También puede otorgarles a los usuarios credenciales de seguridad temporales para limitar su acceso a los datos cargados. Las credenciales de seguridad temporales proporcionan mayor seguridad debido a su breve vigencia y al hecho de que no se pueden reutilizar cuando vencen.

Amazon Redshift tiene características integradas en COPY para cargar datos delimitados y sin comprimir de forma rápida. No obstante, puede comprimir sus archivos utilizando los formatos gzip, lzop o bzip2 para ahorrar tiempo cuando cargue los archivos.

No se admitirá la división automática de datos sin comprimir si las siguientes palabras clave aparecen en la consulta COPY: ESCAPE, REMOVEQUOTES y FIXEDWIDTH. Pero sí se admite la palabra clave CSV.

Para ayudar a mantener la protección de los datos en tránsito en la nube de AWS, Amazon Redshift utiliza la tecnología SSL con aceleración por hardware para comunicarse con Amazon S3 o Amazon DynamoDB en las operaciones de COPY, UNLOAD, copia de seguridad y restauración.

Si carga una tabla directamente desde una tabla de Amazon DynamoDB, puede optar por controlar la cantidad de rendimiento aprovisionado por Amazon DynamoDB que se va a consumir.

De forma alternativa, puede permitirle a COPY que analice los datos de entrada y aplique las codificaciones de compresión óptimas a su tabla de manera automática, como parte del proceso de carga.

Temas


- Credenciales y permisos de acceso
- Preparación de los datos de entrada
- Carga de datos desde Amazon S3
- Carga de datos desde Amazon EMR
- Carga de datos desde hosts remotos
- Carga de datos desde una tabla de Amazon DynamoDB
- Comprobación de carga correcta de datos
- Validación de datos de entrada
- Carga de tablas con compresión automática
- Optimización del almacenamiento para tablas angostas
- Carga de valores de columna predeterminados
- Solución de problemas en cargas de datos


## Credenciales y permisos de acceso

Para cargar o descargar datos con otro recurso de AWS, como Amazon S3, Amazon DynamoDB, Amazon EMR o Amazon EC2, Amazon Redshift debe tener permiso para acceder al recurso y realizar las acciones necesarias para acceder a los datos. Por ejemplo, para cargar datos de Amazon S3, COPY debe tener acceso LIST para el bucket y acceso GET para los objetos del bucket.

Para obtener una autorización para acceder a un recurso, Amazon Redshift debe estar autenticado. Puede elegir un control de acceso basado en un rol o en una clave. En esta sección, se presenta información general relacionada con los dos métodos. Para obtener todos los detalles y ejemplos, consulte Permisos para acceder a otros recursos de AWS.

## Control de acceso con base en roles

Con el control de acceso basado en roles, Amazon Redshift adopta de forma temporal un rol de AWS Identity and Access Management (IAM) en su nombre. Luego, en función de las autorizaciones que se otorgaron al rol, Amazon Redshift puede acceder a los recursos de AWS requeridos.

Recomendamos el uso del control de acceso basado en roles porque brinda un control más preciso y seguro del acceso a los recursos de AWS y a la información confidencial de los usuarios, además de salvaguardar las credenciales de AWS.

Para utilizar el control de acceso basado en roles, primero debe crear un rol de IAM mediante el tipo de rol de servicio de Amazon Redshift y, luego, asociar el rol al almacenamiento de datos. Como mínimo, el rol debe tener los permisos especificados en Permisos de IAM para COPY, UNLOAD y CREATE LIBRARY. Para obtener información acerca de cómo crear un rol de IAM y adjuntarlo al clúster, consulte Creación de un rol de IAM que permita al clúster de Amazon Redshift acceder a los servicios de AWS en la Guía de administración de Amazon Redshift.

Puede agregar un rol a un clúster o ver los roles asociados a un clúster mediante la consola de administración, la CLI o la API de Amazon Redshift. Para obtener más información, consulte Autorización del uso de las operaciones COPY y UNLOAD mediante roles de IAM en la Guía de administración de Amazon Redshift.

Al crear un rol de IAM, IAM devuelve el Amazon Resource Name (ARN, Nombre de recurso de Amazon) del rol. Para ejecutar un comando COPY con un rol de IAM, proporcione el ARN del rol mediante el parámetro IAM_ROLE o el parámetro CREDENTIALS.

En el siguiente ejemplo de comando COPY, se usa el parámetro IAM_ROLE con el rol `MyRedshiftRole` para su autenticación.

```
COPY customer FROM 's3://amzn-s3-demo-bucket/mydata'
IAM_ROLE 'arn:aws:iam::12345678901:role/MyRedshiftRole';
```

Como mínimo, el usuario de AWS debe tener los permisos que aparecen en Permisos de IAM para COPY, UNLOAD y CREATE LIBRARY.

## Control de acceso basado en claves

Con el control de acceso basado en claves, se proporciona el ID de clave de acceso y la clave de acceso secreta de un usuario que está autorizado a acceder a los recursos de AWS que contienen los datos.

> **Note**
>
> Recomendamos encarecidamente que use un rol de IAM para su autenticación, en lugar de brindar un ID de clave de acceso y una clave de acceso secreta como texto sin formato. Si elige el control de acceso basado en claves, nunca utilice las credenciales (raíz) de su cuenta de AWS. Siempre cree un usuario de IAM y proporcione el ID de clave de acceso y la clave de acceso secreta de ese usuario. Si desea conocer los pasos necesarios para crear un usuario de IAM, consulte Creación de un usuario de IAM en su cuenta de AWS.

## Preparación de los datos de entrada

Si los datos de entrada no son compatibles con las columnas de la tabla que los recibirá, el comando COPY dará error.

Use las siguientes directrices para ayudar a garantizar que los datos de entrada sean válidos:

- Los datos solo pueden contener caracteres multibyte UTF-8 de hasta un máximo de cuatro bytes.
- Compruebe que las cadenas CHAR y VARCHAR no sean más largas que las columnas correspondientes. Las cadenas VARCHAR se miden en bytes, no en caracteres, por lo que, por ejemplo, una cadena de cuatro caracteres chinos que ocupan cuatro bytes cada uno requiere una columna VARCHAR(16).
- Los caracteres multibyte solo pueden usarse en columnas VARCHAR. Compruebe que los caracteres multibyte no sean de más de cuatro bytes.
- Compruebe que los datos para las columnas CHAR solo contengan caracteres de un byte.
- No incluya sintaxis ni caracteres especiales para indicar el último campo de un registro. Este campo puede ser un delimitador.
- Si sus datos incluyen terminaciones nulas, también denominadas NUL (UTF-8 0000) o cero binario (0x000), puede cargar esos caracteres como NULLS en columnas CHAR o VARCHAR mediante la opción NULL AS del comando COPY: `null as '\0'` o `null as '\000'`. Si no usa NULL AS, las terminaciones nulas harán que COPY produzca un error.
- Si sus cadenas tienen caracteres especiales, como delimitadores y caracteres de línea nueva insertados, use la opción ESCAPE del comando COPY.
- Compruebe que todas las comillas simples y dobles estén en pares correctamente.
- Compruebe que las cadenas de punto flotante estén en formato de punto flotante estándar, como 12.123, o en formato exponencial, como 1.0E4.

- Compruebe que todas las cadenas de fecha y marca temporal sigan las especificaciones para Cadenas TIMEFORMAT y DATEFORMAT. El formato de marca temporal predeterminado es AAAA-MM-DD hh:mm:ss y el formato de fecha predeterminado es AAAA-MM-DD.
- Para obtener más información acerca de los límites y las limitaciones de los tipos de datos individuales, consulte Tipos de datos. Para obtener más información acerca de errores con caracteres multibyte, consulte Errores de carga de caracteres multibyte.

## Carga de datos desde Amazon S3

El comando COPY utiliza la arquitectura de procesamiento masivo en paralelo (MPP) de Amazon Redshift para leer y cargar datos en paralelo desde uno o varios archivos de un bucket de Amazon S3. Puede aprovechar al máximo los beneficios del procesamiento en paralelo mediante la división de los datos en distintos archivos, en los casos en que los archivos estén comprimidos. (Hay excepciones a esta regla. Se detallan en Divida los datos de carga). También puede aprovechar al máximo los beneficios del procesamiento en paralelo mediante la configuración de claves de distribución en las tablas. Para obtener más información acerca de las claves de distribución, consulte Distribución de datos para la optimización de consultas.

Los datos se cargan en la tabla destino, una línea por fila. Los campos en el archivo de datos se corresponden con las columnas de la tabla, de izquierda a derecha. Los campos en los archivos de datos pueden ser de ancho fijo o con caracteres delimitados; el delimitador predeterminado es la barra vertical (|). De manera predeterminada, se cargan todas las columnas de la tabla, pero de manera alternativa se puede definir una lista de columnas separadas por comas. Si una columna de la tabla no se encuentra en la lista de columnas especificada en el comando COPY, se carga con un valor predeterminado. Para obtener más información, consulte Carga de valores de columna predeterminados.

Temas

- Carga de datos desde archivos comprimidos y sin comprimir
- Carga de archivos en Amazon S3 para utilizar con COPY
- Uso del comando COPY para cargar desde Amazon S3

### Carga de datos desde archivos comprimidos y sin comprimir

Cuando cargue datos comprimidos, le recomendamos que divida los datos de cada tabla en distintos archivos. Cuando carga datos delimitados y sin comprimir, el comando COPY utiliza rangos de

escaneo y procesamiento masivo en paralelo (MPP) para cargar datos de archivos de gran tamaño en un bucket de Amazon S3.

Carga de datos desde varios archivos comprimidos

Cuando tenga datos comprimidos, le recomendamos que divida los datos de cada tabla en varios archivos. El comando COPY puede cargar datos desde distintos archivos en paralelo. Puede cargar varios archivos especificando un prefijo común o el prefijo de clave del conjunto, o enumerando explícitamente los archivos en un archivo de manifiesto.

Divida los datos en archivos de manera tal que la cantidad de archivos sea múltiplo de la cantidad de sectores en su clúster. De esa manera, Amazon Redshift puede dividir los datos de forma uniforme entre los sectores. El número de sectores por nodo depende del tamaño de nodo del clúster. Por ejemplo, cada nodo de computación dc2.large tiene dos sectores y cada nodo de computación dc2.8xlarge tiene 16 sectores. Para obtener más información acerca de la cantidad de sectores que tiene cada tamaño de nodo, consulte Acerca de clústeres y nodos en la Guía de administración de Amazon Redshift.

Todos los nodos participan en la ejecución de consultas en paralelo y trabajan en datos que se distribuyen de la manera más uniforme posible entre los sectores. Si tiene un clúster con dos nodos dc2.large, podría dividir los datos en cuatro archivos o en un múltiplo de cuatro. Amazon Redshift no tiene en cuenta el tamaño de los archivos a la hora de dividir la carga de trabajo. Por lo tanto, debe asegurarse de que los archivos tengan aproximadamente el mismo tamaño, de 1 MB a 1 GB después de la compresión.

Para utilizar prefijos de objeto para identificar los archivos de carga, nombre cada archivo con un prefijo común. Por ejemplo, si desea dividir el archivo `venue.txt` podría hacerlo en cuatro archivos, de la siguiente manera:

```
venue.txt.1
venue.txt.2
venue.txt.3
venue.txt.4
```

Si coloca en su bucket una carpeta con archivos distintos, y nombra a esa carpeta como el prefijo, COPY cargará todos los archivos que se encuentren en ella. Si detalla de manera explícita los archivos que deben cargarse mediante un archivo de manifiesto, los archivos pueden encontrarse en diferentes buckets o carpetas.

Para obtener más información sobre los archivos de manifiesto, consulte Example: COPY from Amazon S3 using a manifest.

Carga de datos desde archivos delimitados y sin comprimir

Cuando se cargan datos delimitados y sin comprimir, el comando COPY utiliza la arquitectura de procesamiento masivo en paralelo (MPP), de Amazon Redshift. Amazon Redshift utiliza automáticamente sectores que funcionan en paralelo para cargar rangos de datos de un archivo de gran tamaño en un bucket de Amazon S3. El archivo debe estar delimitado para que se produzca una carga paralela. Por ejemplo, delimitado por canalización. La carga automática de datos en paralelo con el comando COPY también está disponible para los archivos CSV. Igualmente, puede aprovechar los beneficios del procesamiento en paralelo mediante la configuración de claves de distribución en las tablas. Para obtener más información acerca de las claves de distribución, consulte Distribución de datos para la optimización de consultas.

No se admite la carga automática de datos en paralelo cuando la consulta COPY incluye cualquiera de las siguientes palabras clave: ESCAPE, REMOVEQUOTES y FIXEDWIDTH.

Los datos de uno o más archivos se cargan en la tabla destino, una línea por fila. Los campos en el archivo de datos se corresponden con las columnas de la tabla, de izquierda a derecha. Los campos en los archivos de datos pueden ser de ancho fijo o con caracteres delimitados; el delimitador predeterminado es la barra vertical (|). De manera predeterminada, se cargan todas las columnas de la tabla, pero de manera alternativa se puede definir una lista de columnas separadas por comas. Si una columna de la tabla no está incluida en la lista de columnas especificada en el comando COPY, esta se cargará con un valor predeterminado. Para obtener más información, consulte Carga de valores de columna predeterminados.

Siga este proceso general para cargar datos de Amazon S3, cuando los datos estén delimitados y sin comprimir:

1. Cargue sus archivos en Amazon S3.
2. Ejecute el comando COPY para cargar la tabla.
3. Compruebe que los datos se hayan cargado correctamente.

Para ver ejemplos de comandos COPY, consulte Ejemplos de COPY. Para obtener información acerca de los datos cargados en Amazon Redshift, consulte las tablas de sistema STL_LOAD_COMMITS y STL_LOAD_ERRORS.

Para obtener más información acerca de los nodos y de los sectores que tiene cada nodo, consulte Acerca de clústeres y nodos en la Guía de administración de Amazon Redshift.

## Carga de archivos en Amazon S3 para utilizar con COPY

Hay un par de enfoques que se deben adoptar cuando se carguen archivos de texto en Amazon S3:

- Si tiene archivos comprimidos, le recomendamos que divida los archivos de gran tamaño para aprovechar los beneficios del procesamiento en paralelo en Amazon Redshift.
- Por otro lado, COPY divide automáticamente los datos de archivos de texto de gran tamaño que estén delimitados y sin comprimir para facilitar el paralelismo y distribuir eficazmente los datos de archivos de gran tamaño.

Cree un bucket de Amazon S3 para que contenga sus archivos de datos y, luego, cargue los archivos de datos en él. Para obtener más información acerca de la creación de buckets y la carga de archivos, consulte Trabajo con buckets de Amazon S3 en la Guía del usuario de Amazon Simple Storage Service.

> **Important**
>
> El bucket de Amazon S3 que tiene los archivos de datos debe crearse en la misma región de AWS que el clúster, a menos que use la opción REGION para especificar la región en la que se encuentra el bucket de Amazon S3.

Asegúrese de que los rangos de IP de S3 estén agregados a la lista de permitidos. Para obtener más información acerca de los rangos de IP de S3 necesarios, consulte Aislamiento de red.

Puede crear un bucket de Amazon S3 en una región específica. Para ello, seleccione la región mientras crea el bucket con la consola de Amazon S3 o especifique un punto de conexión mientras crea el bucket con la CLI o la API de Amazon S3.

Luego de la carga de los datos, corrobore que los archivos presentes en Amazon S3 sean los correctos.

Temas

- Administración de la consistencia de los datos
- Carga de datos cifrados en Amazon S3

- Comprobación de que los archivos presentes en el bucket son los correctos

### Administración de la consistencia de los datos

Amazon S3 proporciona una sólida consistencia de lectura tras escritura para las operaciones COPY, UNLOAD, INSERT (tabla externa), CREATE EXTERNAL TABLE AS y Amazon Redshift Spectrum en los buckets de Amazon S3 en todas las regiones de AWS. Además, las operaciones de lectura en Amazon S3 Select, las listas de control de acceso de Amazon S3, las etiquetas de objeto de Amazon S3 y los metadatos de objetos (por ejemplo, el objeto HEAD) son muy consistentes. Para obtener más información acerca de la consistencia de los datos, consulte Modelo de coherencia de datos de Amazon S3 en la Guía del usuario de Amazon Simple Storage Service.

### Carga de datos cifrados en Amazon S3

Amazon S3 admite tanto el cifrado del lado del servidor como el cifrado del lado del cliente. En este tema, se analizan las diferencias entre el cifrado del lado del cliente y el cifrado del lado del servidor, y se describen los pasos para usar un cifrado del lado del cliente con Amazon Redshift. El cifrado del lado del servidor es transparente para Amazon Redshift.

#### Cifrado en el servidor

El cifrado del lado del servidor es el cifrado de datos en reposo; es decir, Amazon S3 cifra los datos a medida que los carga y los descifra para usted cuando obtiene acceso a estos. Cuando se cargan tablas con un comando COPY, no existe diferencia alguna en la forma de carga desde objetos con cifrado del lado del servidor o sin cifrado en Amazon S3. Para obtener más información acerca del cifrado del lado del servidor, consulte Uso del cifrado del lado del servidor en la Guía del usuario de Amazon Simple Storage Service.

#### Cifrado del cliente

En el cifrado en el cliente, su aplicación cliente administra el cifrado de los datos, las claves de cifrado y las herramientas relacionadas. Puede cargar los datos a un bucket de Amazon S3 con el cifrado del lado del cliente y, luego, cargar los datos con el comando COPY con la opción ENCRYPTED y una clave de cifrado privada para proporcionar mayor seguridad.

Usted cifra los datos con cifrado de sobre. Con el cifrado de sobre, la aplicación controla todo el cifrado de manera exclusiva. Las claves de cifrado privadas y los datos sin cifrar nunca se envían a AWS, por lo que es muy importante que administre con seguridad las claves de cifrado. Si pierde las claves de cifrado, no podrá descifrar los datos, y estas no se podrán recuperar desde AWS.

El cifrado de sobre combina el rendimiento de un cifrado simétrico rápido al mismo tiempo que mantiene la mayor seguridad que brinda la administración de claves con claves asimétricas. El cliente de cifrado de Amazon S3 genera una clave simétrica de un solo uso (la clave simétrica de sobre) para cifrar los datos. Luego, esa clave se cifra mediante la clave raíz y se almacena junto con los datos en Amazon S3. Cuando Amazon Redshift obtiene acceso a los datos durante la carga, la clave simétrica cifrada se recupera y se descifra con la clave real y, luego, se descifran los datos.

Para trabajar en Amazon Redshift con datos cifrados del lado del cliente de Amazon S3, siga los pasos que se indican en el artículo Protección de datos con el cifrado del lado del cliente en la Guía del usuario de Amazon Simple Storage Service, con el requisito adicional de que utilice lo siguiente:

- Cifrado simétrico: el AWS SDK para Java en su clase `AmazonS3EncryptionClient` usa el cifrado de sobre, descrito antes, que se basa en el cifrado con claves simétricas. Use esta clase para crear un cliente de Amazon S3 en el que cargar datos cifrados del lado del cliente.
- Una clave raíz simétrica con cifrado AES de 256 bits: una clave raíz cifra la clave de sobre. La clave raíz se pasa a la instancia de la clase `AmazonS3EncryptionClient`. Guarde esa clave, ya que la necesitará para copiar los datos en Amazon Redshift.
- Metadatos de los objetos que van a almacenar la clave de sobre cifrada: de forma predeterminada, Amazon S3 almacena la clave de sobre como metadatos de los objetos de la clase `AmazonS3EncryptionClient`. La clave de sobre cifrada que se almacena como metadatos de los objetos se usa durante el proceso de descifrado.

> **Note**
>
> Si obtiene un mensaje de error de cifrado cuando usa la Application Programming Interface (API, Interfaz de programación de aplicaciones) de cifrado por primera vez, su versión del Java Development Kit (JDK, Kit de desarrollo de Java) puede tener un archivo de política de jurisdicción de Java Cryptography Extension (JCE, Extensión de criptografía Java) que limite la longitud máxima de la clave para las transformaciones de cifrado y descifrado a 128 bits. Para obtener información acerca de cómo solucionar este problema, consulte Especificación del cifrado del lado del cliente con AWS SDK para Java en la Guía del usuario de Amazon Simple Storage Service.

Para obtener información acerca de la carga de archivos cifrados del lado del cliente en las tablas de Amazon Redshift con el comando COPY, consulte Carga de archivos de datos cifrados desde Amazon S3.

Ejemplo: Carga de datos cifrados en el cliente

Si desea obtener un ejemplo sobre cómo se utiliza AWS SDK para Java para cargar datos cifrados del lado del cliente, consulte Protección de datos mediante el cifrado del lado del cliente en la Guía del usuario de Amazon Simple Storage Service.

En la segunda opción, se muestran las decisiones que debe tomar durante el cifrado del lado del cliente para que los datos puedan cargarse en Amazon Redshift. Específicamente, en el ejemplo se muestra el uso de metadatos de los objetos para almacenar la clave de sobre cifrada y el uso de una clave raíz simétrica con cifrado AES de 256 bits.

En este ejemplo, se proporciona un código de ejemplo que utiliza AWS SDK para Java para crear una clave raíz simétrica con cifrado AES de 256 bits y guardarla en un archivo. Luego, se incluye un ejemplo de carga de un objeto en Amazon S3 con un cliente de cifrado de S3 que primero cifra los datos de muestra en el cliente. En el ejemplo, también se descarga el objeto y se comprueba que los datos sean los mismos.

Comprobación de que los archivos presentes en el bucket son los correctos

Una vez que cargue los archivos en su bucket de Amazon S3, le recomendamos que enumere el contenido del bucket para corroborar que este tenga todos los archivos correctos y que no incluya ningún archivo no deseado. Por ejemplo, si el bucket `amzn-s3-demo-bucket` tiene un archivo denominado `venue.txt.back`, ese archivo será cargado, quizás de manera no intencional, por el siguiente comando:

```
COPY venue FROM 's3://amzn-s3-demo-bucket/venue' … ;
```

Si desea controlar de manera específica qué archivos se cargan, puede usar un archivo de manifiesto para listar los archivos de datos de manera explícita. Para obtener más información acerca del uso de un archivo de manifiesto, consulte la opción copy_from_s3_manifest_file para el comando COPY y Example: COPY from Amazon S3 using a manifest en los ejemplos de COPY.

Para obtener más información acerca de cómo mostrar el contenido del bucket, consulte Listas de claves de objetos en la Guía para desarrolladores de Amazon S3.

## Uso del comando COPY para cargar desde Amazon S3

Uso del comando COPY para cargar una tabla en paralelo desde los archivos de datos en Amazon S3. Puede especificar los archivos que se cargarán mediante un prefijo de objeto de Amazon S3 o un archivo de manifiesto.

La sintaxis para especificar los archivos por cargar mediante el uso de un prefijo es la siguiente:

```
COPY <table_name> FROM 's3://<bucket_name>/<object_prefix>'
authorization;
```

El archivo de manifiesto es un archivo con formato JSON que lista los archivos de datos que se deben cargar. La sintaxis para especificar los archivos por cargar mediante el uso de un archivo de manifiesto es la siguiente:

```
COPY <table_name> FROM 's3://<bucket_name>/<manifest_file>'
authorization
MANIFEST;
```

La tabla que se cargará ya debe existir en la base de datos. Para obtener más información acerca de la creación de una tabla, consulte CREATE TABLE en la referencia de SQL.

Los valores de autorización proporcionan la autorización de AWS que Amazon Redshift necesita para acceder a los objetos de Amazon S3. Para obtener más información acerca de los permisos necesarios, consulte Permisos de IAM para COPY, UNLOAD y CREATE LIBRARY. El método de autenticación preferido es especificar el parámetro IAM_ROLE y proporcionar el Amazon Resource Name (ARN, Nombre de recurso de Amazon) de un rol de IAM con los permisos necesarios. Para obtener más información, consulte Control de acceso con base en roles.

Para realizar la autenticación con el parámetro IAM_ROLE, reemplace *<aws-account-id>* y *<role-name>*, tal y como se observa en la siguiente sintaxis.

```
IAM_ROLE 'arn:aws:iam::<aws-account-id>:role/<role-name>'
```

En el siguiente ejemplo, se muestra la autenticación con un rol de IAM.

```
COPY customer
FROM 's3://amzn-s3-demo-bucket/mydata'
IAM_ROLE 'arn:aws:iam::0123456789012:role/MyRedshiftRole';
```

Para obtener más información acerca de otras opciones de autorización, consulte Parámetros de autorización.

Si desea validar los datos sin cargar de hecho la tabla, use la opción NOLOAD con el comando COPY.

En el siguiente ejemplo, se muestran las primeras filas de datos delimitados por barras verticales en un archivo denominado `venue.txt`.

```
1|Toyota Park|Bridgeview|IL|0
2|Columbus Crew Stadium|Columbus|OH|0
3|RFK Stadium|Washington|DC|0
```

Antes de cargar el archivo en Amazon S3, divídalo en distintos archivos para que el comando COPY pueda cargarlo usando el procesamiento en paralelo. La cantidad de archivos debe ser múltiplo de la cantidad de secciones en su clúster. Divida los archivos de datos de carga de modo tal que sean de igual tamaño, entre 1 MB y 1 GB, después de la compresión. Para obtener más información, consulte Carga de datos desde archivos comprimidos y sin comprimir.

Por ejemplo, el archivo `venue.txt` podría dividirse en cuatro archivos, de la siguiente manera:

```
venue.txt.1
venue.txt.2
venue.txt.3
venue.txt.4
```

El siguiente comando COPY carga la tabla VENUE con datos delimitados por barras verticales en los archivos de datos con el prefijo “venue” del bucket `amzn-s3-demo-bucket` de Amazon S3.

> **Note**
>
> El bucket `amzn-s3-demo-bucket` de Amazon S3 no está presente en los siguientes ejemplos. Para ver información acerca de comandos COPY de muestra que usan datos reales de un bucket de Amazon S3 existente, consulte Carga de datos de muestra.

```
COPY venue FROM 's3://amzn-s3-demo-bucket/venue'
IAM_ROLE 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
DELIMITER '|';
```

Si no hay objetos de Amazon S3 con el prefijo de clave “venue”, se producirá un error en la carga.

Temas

- Uso de un manifiesto para especificar archivos de datos

- Carga de archivos de datos comprimidos desde Amazon S3
- Carga de datos de ancho fijo desde Amazon S3
- Carga de datos multibyte desde Amazon S3
- Carga de archivos de datos cifrados desde Amazon S3

Uso de un manifiesto para especificar archivos de datos

Puede usar un manifiesto para asegurarse de que el comando COPY cargue todos los archivos requeridos, y solo los requeridos, en la carga de datos. Puede usar un manifiesto para cargar archivos de diferentes buckets o archivos que no comparten el mismo prefijo. En lugar de proporcionar la ruta de un objeto para el comando COPY, se proporciona el nombre de un archivo de texto en formato JSON que lista explícitamente los archivos por cargar. El URL en el manifiesto debe especificar el nombre del bucket y la ruta completa del objeto para el archivo, no solo un prefijo.

Para obtener más información acerca de los archivos de manifiesto, consulte el ejemplo COPY Uso de un manifiesto para especificar archivos de datos.

En el siguiente ejemplo, se muestra el JSON para cargar archivos de diferentes buckets y con nombres de archivo que comiencen con marcas de fecha.

```
{
  "entries": [
    {"url":"s3://amzn-s3-demo-bucket1/2013-10-04-custdata", "mandatory":true},
    {"url":"s3://amzn-s3-demo-bucket1/2013-10-05-custdata", "mandatory":true},
    {"url":"s3://amzn-s3-demo-bucket2/2013-10-04-custdata", "mandatory":true},
    {"url":"s3://amzn-s3-demo-bucket2/2013-10-05-custdata", "mandatory":true}
  ]
}
```

La marca opcional `mandatory` especifica si COPY debe devolver un error en caso de que el archivo no se encuentre. El valor predeterminado de `mandatory` es `false`. Independientemente de cualquier configuración obligatoria, COPY terminará si no se encuentran archivos.

En el siguiente ejemplo, se ejecuta el comando COPY con el manifiesto del ejemplo anterior, que se denomina `cust.manifest`.

```
COPY customer
FROM 's3://amzn-s3-demo-bucket/cust.manifest'
```

```
IAM_ROLE 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
MANIFEST;
```

Uso de un manifiesto creado por UNLOAD

Un manifiesto creado por una operación UNLOAD que usa el parámetro MANIFEST podría tener claves no requeridas para la operación COPY. Por ejemplo, en el siguiente manifiesto de UNLOAD, se incluye una clave `meta` que es necesaria para una tabla externa de Amazon Redshift Spectrum y para cargar archivos de datos con el formato `ORC` o `Parquet`. La clave `meta` contiene una clave `content_length` con un valor que es el tamaño real del archivo en bytes. La operación COPY requiere solo la clave `url` y una clave `mandatory` opcional.

```
{
  "entries": [
    {"url":"s3://amzn-s3-demo-bucket/unload/manifest_0000_part_00", "meta":
 { "content_length": 5956875 }},
    {"url":"s3://amzn-s3-demo-bucket/unload/unload/manifest_0001_part_00", "meta":
 { "content_length": 5997091 }}
 ]
}
```

Para obtener más información sobre los archivos de manifiesto, consulte Example: COPY from Amazon S3 using a manifest.

Carga de archivos de datos comprimidos desde Amazon S3

Para cargar archivos de datos comprimidos con gzip, lzop o bzip2, incluya la opción que corresponda: GZIP, LZOP o BZIP2.

Por ejemplo, el siguiente comando carga desde archivos comprimidos con lzop.

```
COPY customer FROM 's3://amzn-s3-demo-bucket/customer.lzo'
IAM_ROLE 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
DELIMITER '|' LZOP;
```

> **Note**
>
> Si comprime un archivo de datos con la compresión lzop y utiliza la opción --filter, el comando COPY no la admite.

## Carga de datos de ancho fijo desde Amazon S3

Los archivos de datos de ancho fijo tienen longitudes uniformes para cada columna de datos. Cada campo de un archivo con datos de ancho fijo tiene exactamente la misma longitud y posición. En el caso de datos de caracteres (CHAR y VARCHAR) en un archivo de datos de ancho fijo, debe incluir espacios a la izquierda o a la derecha como marcadores de posición, con objeto de mantener la uniformidad del ancho. En el caso de números enteros, se deben usar ceros a la izquierda como marcadores de posición. Un archivo de datos de ancho fijo no tiene delimitador para separar columnas.

Para cargar un archivo de datos de ancho fijo en una tabla existente, USE el parámetro FIXEDWIDTH del comando COPY. Las especificaciones de la tabla deben coincidir con el valor de fixedwidth_spec con objeto de que los datos se carguen de manera correcta.

Para cargar datos de ancho fijo en una tabla desde un archivo, emita el siguiente comando:

```
COPY table_name FROM 's3://amzn-s3-demo-bucket/prefix'
IAM_ROLE 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
FIXEDWIDTH 'fixedwidth_spec';
```

El parámetro fixedwidth_spec es una cadena que contiene un identificador para cada columna y el ancho de cada columna, separados por dos puntos. Los pares **`column:width`** están delimitados por comas. El identificador puede ser lo que usted elija: números, letras o una combinación de los dos. El identificador no se relaciona con la tabla en sí, por lo que la especificación debe tener las columnas en el mismo orden que la tabla.

En los siguientes dos ejemplos, se observa la misma especificación; la primera usa identificadores numéricos y la segunda usa identificadores de cadena:

```
'0:3,1:25,2:12,3:2,4:6'
```

```
'venueid:3,venuename:25,venuecity:12,venuestate:2,venueseats:6'
```

En el siguiente ejemplo, se muestran datos de muestra de ancho fijo que podrían cargarse en la tabla VENUE con las especificaciones anteriores:

```
1  Toyota Park              Bridgeview  IL0
2  Columbus Crew Stadium    Columbus    OH0
3  RFK Stadium              Washington  DC0
```

```
4  CommunityAmerica Ballpark Kansas City KS0
5  Gillette Stadium          Foxborough  MA68756
```

El siguiente comando COPY carga este conjunto de datos en la tabla VENUE:

```
COPY venue
FROM 's3://amzn-s3-demo-bucket/data/venue_fw.txt'
IAM_ROLE 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
FIXEDWIDTH 'venueid:3,venuename:25,venuecity:12,venuestate:2,venueseats:6';
```

Carga de datos multibyte desde Amazon S3

Si los datos incluyen caracteres multibyte no ASCII (como caracteres cirílicos o chinos) debe cargar los datos en columnas VARCHAR. El tipo de datos VARCHAR admite caracteres UTF-8 de cuatro bytes, pero el tipo de datos CHAR solo acepta caracteres ASCII de un solo byte. No puede cargar caracteres de cinco bytes o más en tablas de Amazon Redshift. Para obtener más información acerca de los comandos CHAR y VARCHAR, consulte Tipos de datos.

Para comprobar qué cifrado usa un archivo de entrada, use el comando *file* de Linux:

```
$ file ordersdata.txt
ordersdata.txt: ASCII English text
$ file uni_ordersdata.dat
uni_ordersdata.dat: UTF-8 Unicode text
```

Carga de archivos de datos cifrados desde Amazon S3

Puede usar el comando COPY para cargar archivos de datos cargados en Amazon S3 usando el cifrado del lado del servidor, el cifrado del lado del cliente o de ambos.

El comando COPY es compatible con los siguientes tipos de cifrado de Amazon S3:

- Cifrado en el servidor con claves administradas por Amazon S3 (SSE-S3)
- Cifrado del lado del servidor con AWS KMS keys (SSE-KMS)
- Cifrado del lado del cliente con una clave raíz simétrica del lado del cliente

El comando COPY no es compatible con los siguientes tipos de cifrado de Amazon S3:

- Cifrado en el servidor con claves proporcionadas por el cliente (SSE-C)
- Cifrado del lado del cliente con una AWS KMS key

- Cifrado del lado del cliente con una clave raíz asimétrica provista por el cliente

Para obtener más información acerca del cifrado de Amazon S3, consulte Protección de datos con el cifrado del lado del servidor y Protección de datos con el cifrado del lado del cliente en la Guía del usuario de Amazon Simple Storage Service.

El comando UNLOAD cifra los archivos con SSE-S3 de manera automática. También puede descargar con cifrado del cliente o SSE-KMS con una clave simétrica administrada por el cliente. Para obtener más información, consulte Descarga de archivos de datos cifrados

El comando COPY reconoce automáticamente y carga los archivos cifrados con SSE-S3 y SSE-KMS. Puede cargar archivos cifrados con una clave raíz simétrica del lado del cliente especificando la opción ENCRYPTED y proporcionando el valor de clave. Para obtener más información, consulte Carga de datos cifrados en Amazon S3.

Para cargar archivos de datos cifrados del lado del cliente, proporcione el valor de clave raíz con el parámetro MASTER_SYMMETRIC_KEY e incluya la opción ENCRYPTED.

```
COPY customer FROM 's3://amzn-s3-demo-bucket/encrypted/customer'
IAM_ROLE 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
MASTER_SYMMETRIC_KEY '<root_key>'
ENCRYPTED
DELIMITER '|';
```

Para cargar archivos de datos cifrados que están comprimidos en gzip, lzop o bzip2, incluya la opción GZIP, LZOP o BZIP2 junto con el valor de clave raíz y la opción ENCRYPTED.

```
COPY customer FROM 's3://amzn-s3-demo-bucket/encrypted/customer'
IAM_ROLE 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
MASTER_SYMMETRIC_KEY '<root_key>'
ENCRYPTED
DELIMITER '|'
GZIP;
```

## Carga de datos desde Amazon EMR

Puede utilizar el comando COPY para cargar datos en paralelo desde un clúster de Amazon EMR configurado para escribir archivos de texto en el sistema de archivos distribuido de Hadoop (HDFS) del clúster, como archivos de ancho fijo, archivos delimitados por caracteres, archivos CSV o archivos con formato JSON.

## Proceso para cargar datos desde Amazon EMR

En esta sección, se detalla el proceso de carga de datos desde un clúster de Amazon EMR. En las siguientes secciones, se proporcionan los detalles que debe completar en cada paso.

- Paso 1: Configurar los permisos de IAM

  Los usuarios que crean el clúster de Amazon EMR y ejecutan el comando COPY de Amazon Redshift deben tener los permisos necesarios.

- Paso 2: Crear un clúster de Amazon EMR

  Configure el clúster para que produzca archivos de texto para Hadoop Distributed File System (HDFS). Necesitará el ID de clúster de Amazon EMR y el DNS público principal del clúster (el punto de conexión de la instancia de Amazon EC2 que aloja el clúster).

- Paso 3: Recuperar la clave pública del clúster de Amazon Redshift y las direcciones IP del nodo del clúster

  La clave pública permite a los nodos del clúster de Amazon Redshift establecer conexiones SSH a los alojamientos. Usará la dirección IP para cada nodo del clúster con objeto de configurar los grupos de seguridad del alojamiento para que permitan obtener acceso desde su clúster de Amazon Redshift con esas direcciones IP.

- Paso 4: Agregar la clave pública del clúster de Amazon Redshift al archivo de claves autorizadas de cada alojamiento de Amazon EC2

  Debe agregar la clave pública del clúster de Amazon Redshift al archivo de claves autorizadas del alojamiento para que este último reconozca el clúster de Amazon Redshift y acepte la conexión SSH.

- Paso 5: Configurar los alojamientos para que acepten todas las direcciones IP del clúster de Amazon Redshift

  Modifique los grupos de seguridad de la instancia de Amazon EMR para agregar reglas de entrada que acepten las direcciones IP de Amazon Redshift.

- Paso 6: Ejecutar el comando COPY para cargar los datos

  Ejecute el comando COPY desde una base de datos de Amazon Redshift para cargar los datos en una tabla de Amazon Redshift.

## Paso 1: Configurar los permisos de IAM

Los usuarios que crean el clúster de Amazon EMR y ejecutan el comando COPY de Amazon Redshift deben tener los permisos necesarios.

Pasos para configurar los permisos de IAM

1. Agregue los siguientes permisos para el usuario que creará el clúster de Amazon EMR.

```
ec2:DescribeSecurityGroups
ec2:RevokeSecurityGroupIngress
ec2:AuthorizeSecurityGroupIngress
redshift:DescribeClusters
```

2. Agregue el siguiente permiso para el rol o usuario de IAM que ejecutará el comando COPY.

```
elasticmapreduce:ListInstances
```

3. Agregue el siguiente permiso al rol de IAM del clúster de Amazon EMR.

```
redshift:DescribeClusters
```

## Paso 2: Crear un clúster de Amazon EMR

El comando COPY carga datos de archivos en el Hadoop Distributed File System (HDFS) de Amazon EMR. Cuando cree el clúster de Amazon EMR, configúrelo para que genere archivos de datos para el HDFS del clúster.

Para crear un clúster de Amazon EMR

1. Cree un clúster de Amazon EMR en la misma región de AWS que el clúster de Amazon Redshift.

   Si el clúster de Amazon Redshift se encuentra en una VPC, el clúster de Amazon EMR debe encontrarse en el mismo grupo de la VPC. Si el clúster de Amazon Redshift usa el modo EC2-Classic (es decir, no se encuentra en una VPC), el clúster de Amazon EMR también debe usar el modo EC2-Classic. Para obtener más información, consulte Administración de clústeres en una nube privada virtual (VPC) en la Guía de administración de Amazon Redshift.

2. Configure el clúster para que genere archivos de datos para el HDFS del clúster. Los nombres de los archivos HDFS no deben incluir asteriscos (*) ni signos de interrogación (?).

   > **Important**
   >
   > Los nombres de los archivos no deben incluir asteriscos (*) ni signos de interrogación (?).

3. Especifique No en la opción Auto-terminate (Terminar automáticamente) de la configuración del clúster de Amazon EMR para que el clúster permanezca disponible mientras se ejecuta el comando COPY.

   > **Important**
   >
   > Si se cambia o se elimina cualquiera de los archivos de datos antes de que se complete la operación COPY, se podrían obtener resultados inesperados o la operación COPY podría provocar un error.

4. Tome nota del ID de clúster y del DNS público principal del clúster (el punto de conexión para la instancia de Amazon EC2 que aloja el clúster). Usará esa información en pasos posteriores.

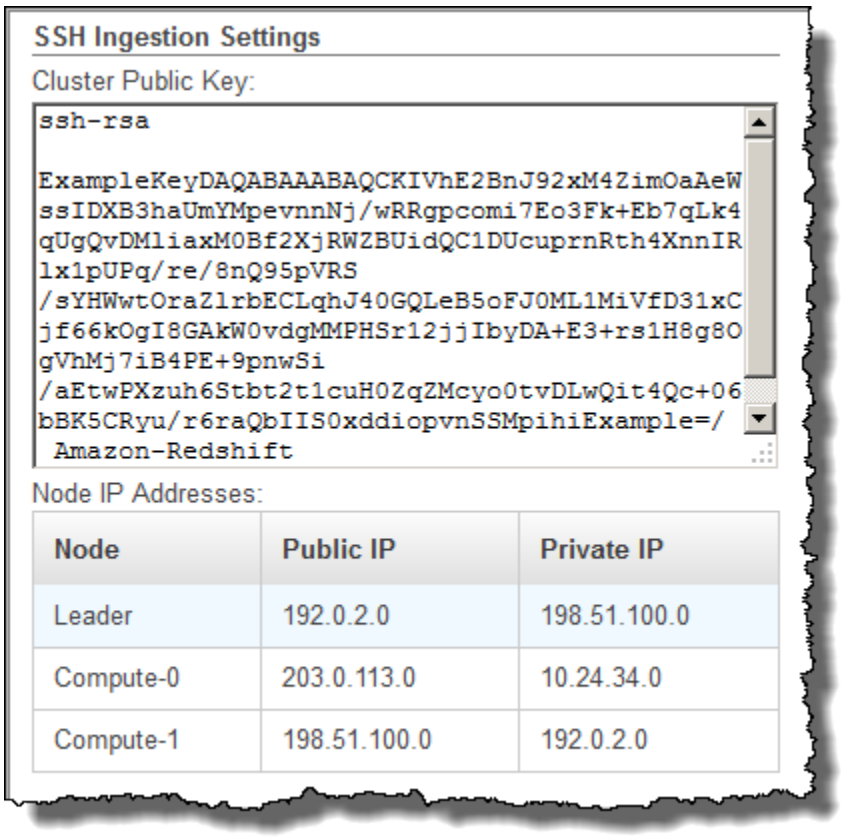
## Paso 3: Recuperar la clave pública del clúster de Amazon Redshift y las direcciones IP del nodo del clúster

Usará la dirección IP para cada nodo del clúster con objeto de configurar los grupos de seguridad del alojamiento para que permitan obtener acceso desde su clúster de Amazon Redshift con esas direcciones IP.

Para recuperar la clave pública del clúster de Amazon Redshift y las direcciones IP del nodo del clúster para su clúster con la consola

1. Acceda a la consola de administración de Amazon Redshift.
2. Elija el enlace de Clusters (Clústeres) en el panel de navegación.
3. Seleccione su clúster de la lista.
4. Localice el grupo SSH Ingestion Settings (Configuración de ingestión de SSH).

   Tome nota de los valores de Cluster Public Key (Clave pública del clúster) y Node IP addresses (Direcciones IP del nodo). Los usará en pasos posteriores.




En el paso 3, usará las direcciones IP privadas para configurar el host de Amazon EC2 de forma que acepte la conexión de Amazon Redshift.

Para recuperar la clave pública del clúster y las direcciones IP del nodo del clúster para su clúster mediante la CLI de Amazon Redshift, ejecute el comando describe-clusters. Por ejemplo:

```
aws redshift describe-clusters --cluster-identifier <cluster-identifier>
```

La respuesta incluirá un valor ClusterPublicKey y la lista de direcciones IP públicas y privadas, similar a la siguiente:

```
{
    "Clusters": [
        {
            "VpcSecurityGroups": [],
            "ClusterStatus": "available",
            "ClusterNodes": [
                {
                    "PrivateIPAddress": "10.nnn.nnn.nnn",
```

```
                    "NodeRole": "LEADER",
                    "PublicIPAddress": "10.nnn.nnn.nnn"
                },
                {
                    "PrivateIPAddress": "10.nnn.nnn.nnn",
                    "NodeRole": "COMPUTE-0",
                    "PublicIPAddress": "10.nnn.nnn.nnn"
                },
                {
                    "PrivateIPAddress": "10.nnn.nnn.nnn",
                    "NodeRole": "COMPUTE-1",
                    "PublicIPAddress": "10.nnn.nnn.nnn"
                }
            ],
            "AutomatedSnapshotRetentionPeriod": 1,
            "PreferredMaintenanceWindow": "wed:05:30-wed:06:00",
            "AvailabilityZone": "us-east-1a",
            "NodeType": "dc2.large",
            "ClusterPublicKey": "ssh-rsa AAAABexamplepublickey...Y3TAl Amazon-
Redshift",
             ...
             ...
}
```

Para recuperar la clave pública del clúster y las direcciones IP del nodo del clúster para su clúster mediante la API de Amazon Redshift, use la acción `DescribeClusters`. Para obtener más información, consulte describe-clusters en la Guía de la CLI de Amazon Redshift o DescribeClusters en la Guía de la API de Amazon Redshift.

## Paso 4: Agregar la clave pública del clúster de Amazon Redshift al archivo de claves autorizadas de cada alojamiento de Amazon EC2

Debe agregar la clave pública del clúster al archivo de claves autorizadas de cada alojamiento en todos los nodos del clúster de Amazon EMR para que los alojamientos reconozcan a Amazon Redshift y acepten la conexión SSH.

Para agregar la clave pública del clúster de Amazon Redshift al archivo de claves autorizadas del alojamiento

1. Obtenga acceso al host con una conexión Secure Shell (SSH, Shell seguro).

Para obtener información acerca de cómo conectarse a una instancia con SSH, consulte Conexión con la instancia en la Guía del usuario de Amazon EC2.

2. Copie la clave pública de Amazon Redshift desde la consola o desde el texto de respuesta de la CLI.
3. Copie y pegue el contenido de la clave pública en el archivo `/home/<ssh_username>/.ssh/authorized_keys` que se encuentra en el host. Incluya la cadena completa, incluso el prefijo "`ssh-rsa`" y el sufijo "`Amazon-Redshift`". Por ejemplo:

```
ssh-rsa AAAACTP3isxgGzVWoIWpbVvRCOzYdVifMrh… uA70BnMHCaMiRdmvsDOedZDOedZ Amazon-
Redshift
```

## Paso 5: Configurar los alojamientos para que acepten todas las direcciones IP del clúster de Amazon Redshift

Para permitir el tráfico de entrada a las instancias del alojamiento, edite el grupo de seguridad y agregue una regla de entrada por cada nodo del clúster de Amazon Redshift. En Type (Tipo), seleccione SSH with TCP protocol on Port 22 (SSH con protocolo TCP en el puerto 22). En Source (Origen), ingrese las direcciones IP privadas de los nodos del clúster de Amazon Redshift que recuperó en Paso 3: Recuperar la clave pública del clúster de Amazon Redshift y las direcciones IP del nodo del clúster. Para obtener más información acerca de cómo se agregan las reglas a un grupo de seguridad de Amazon EC2, consulte Autorización del tráfico de entrada para sus instancias en la Guía del usuario de Amazon EC2.

## Paso 6: Ejecutar el comando COPY para cargar los datos

Ejecute un comando COPY para conectarse al clúster de Amazon EMR y cargar los datos en una tabla de Amazon Redshift. El clúster de Amazon EMR debe seguir ejecutándose hasta que se complete la ejecución del comando COPY. Por ejemplo, no configure el clúster para que se termine automáticamente.

> **Important**
>
> Si se cambia o se elimina cualquiera de los archivos de datos antes de que se complete la operación COPY, se podrían obtener resultados inesperados o la operación COPY podría provocar un error.

En el comando COPY, especifique el ID del clúster de Amazon EMR y el nombre y la ruta del archivo HDFS.

```
COPY sales
FROM 'emr://myemrclusterid/myoutput/part*' CREDENTIALS
IAM_ROLE 'arn:aws:iam::0123456789012:role/MyRedshiftRole';
```

Puede usar los caracteres comodín de asterisco (*) y de signo de interrogación (?) como parte del argumento de nombre del archivo. Por ejemplo, `part*` carga los archivos `part-0000`, `part-0001` y, así, sucesivamente. Si solo especifica el nombre de una carpeta, COPY prueba cargar todos los archivos que se encuentran en ella.

> **Important**
>
> Si usa caracteres comodín o solo el nombre de la carpeta, controle que no se carguen archivos no deseados o el comando COPY provocará un error. Por ejemplo, algunos procesos podrían escribir un archivo de registro en la carpeta de salida.

## Carga de datos desde hosts remotos

Puede utilizar el comando COPY para cargar datos en paralelo desde uno o más alojamientos remotos, como instancias de Amazon EC2 u otros equipos. COPY se conecta a los alojamientos remotos por medio de SSH y ejecuta los comandos en los alojamientos remotos para generar texto de salida.

El alojamiento remoto puede ser una instancia de Linux de Amazon EC2 u otro equipo Linux o Unix configurado para aceptar conexiones SSH. En esta guía, se asume que su alojamiento remoto es una instancia de Amazon EC2. Cuando el procedimiento sea diferente para otro equipo, en la guía se señalará la diferencia.

Amazon Redshift puede conectarse a varios alojamientos y puede establecer varias conexiones SSH en cada alojamiento. Amazon Redshift envía un único comando a través de cada conexión para generar el texto que aparecerá en la salida estándar del alojamiento, que Amazon Redshift leerá después como cualquier otro archivo de texto.

### Antes de empezar

Antes de empezar, debe haber implementado lo siguiente:

- Uno o más equipos de alojamiento, como las instancias de Amazon EC2, a los que puede conectarse mediante SSH.
- Orígenes de datos en los hosts.

  Proporcionará comandos que el clúster de Amazon Redshift ejecutará en los alojamientos para generar el texto de salida. Una vez que el clúster se conecta a un alojamiento, el comando COPY ejecuta los comandos, lee el texto de la salida estándar de los alojamientos y carga los datos en paralelo en una tabla de Amazon Redshift. El texto de salida debe tener un formato que el comando COPY pueda incorporar. Para obtener más información, consulte Preparación de los datos de entrada
- Obtenga acceso a los hosts desde su equipo.

  En caso de una instancia de Amazon EC2, usará una conexión SSH para obtener acceso al alojamiento. Debe acceder al host para agregar la clave pública del clúster de Amazon Redshift al archivo de claves autorizadas del host.
- Un clúster de Amazon Redshift en ejecución.

  Para obtener información acerca de cómo lanzar un clúster, consulte Guía de introducción a Amazon Redshift.

## Proceso de carga de datos

En esta sección, se detalla el proceso de carga de datos desde hosts remotos. En las siguientes secciones, se proporcionan los detalles que debe completar en cada paso.

- Paso 1: Recuperar la clave pública del clúster y las direcciones IP del nodo del clúster

  La clave pública permite a los nodos del clúster de Amazon Redshift establecer conexiones SSH a los alojamientos remotos. Usará la dirección IP para cada nodo del clúster con el fin de configurar el firewall o los grupos de seguridad del alojamiento para que permitan obtener acceso desde su clúster de Amazon Redshift con esas direcciones IP.
- Paso 2: Agregar la clave pública del clúster de Amazon Redshift al archivo de claves autorizadas del alojamiento

  Debe agregar la clave pública del clúster de Amazon Redshift al archivo de claves autorizadas del alojamiento para que este último reconozca el clúster de Amazon Redshift y acepte la conexión SSH.

- Paso 3: Configurar el alojamiento para que acepte todas las direcciones IP del clúster de Amazon Redshift

  Para Amazon EC2, modifique los grupos de seguridad de la instancia para agregar reglas de entrada que acepten las direcciones IP de Amazon Redshift. En caso de otros alojamientos, modifique el firewall de modo que los nodos de Amazon Redshift puedan establecer conexiones SSH al alojamiento remoto.
- Paso 4: Obtener la clave pública para el host

  De manera alternativa, puede especificar que Amazon Redshift debe usar la clave pública para identificar el alojamiento. Debe localizar la clave pública y copiar el texto en el archivo de manifiesto.
- Paso 5: Crear un archivo de manifiesto

  El manifiesto es un archivo de texto con formato JSON que tiene la información que Amazon Redshift necesita para conectarse a los alojamientos y obtener los datos.
- Paso 6: Cargar el archivo de manifiesto en un bucket de Amazon S3

  Amazon Redshift lee el manifiesto y usa esa información para conectarse al alojamiento remoto. Si el bucket de Amazon S3 no se encuentra en la misma región que el clúster de Amazon Redshift, debe usar la opción REGION para especificar la región en la que se encuentran los datos.
- Paso 7: Ejecutar el comando COPY para cargar los datos

  Ejecute el comando COPY desde una base de datos de Amazon Redshift para cargar los datos en una tabla de Amazon Redshift.

## Paso 1: Recuperar la clave pública del clúster y las direcciones IP del nodo del clúster

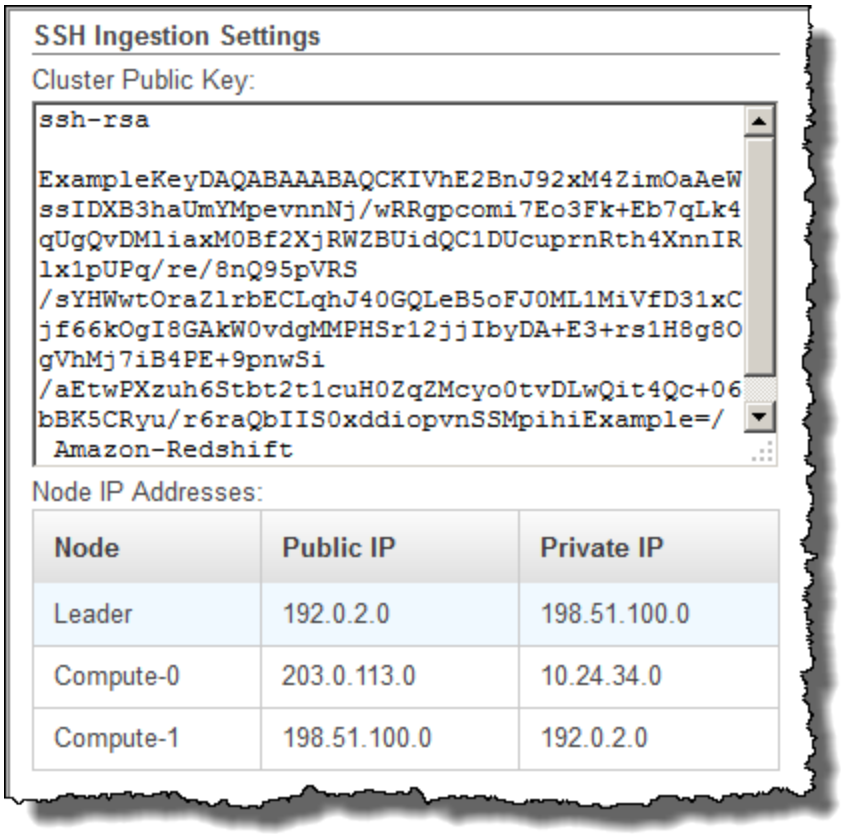
Usará la dirección IP para cada nodo del clúster con objeto de configurar los grupos de seguridad del alojamiento para que permitan obtener acceso desde su clúster de Amazon Redshift con esas direcciones IP.

Pasos para recuperar la clave pública del clúster y las direcciones IP del nodo del clúster para su clúster con la consola

1. Acceda a la consola de administración de Amazon Redshift.
2. Elija el enlace de Clusters (Clústeres) en el panel de navegación.
3. Seleccione su clúster de la lista.

4. Localice el grupo SSH Ingestion Settings (Configuración de ingestión de SSH).

   Tome nota de los valores de Cluster Public Key (Clave pública del clúster) y Node IP addresses (Direcciones IP del nodo). Los usará en pasos posteriores.



SSH Ingestion Settings

Cluster Public Key:

```
ssh-rsa

ExampleKeyDAQABAAABAQCKIVhE2BnJ92xM4ZimOaAeW
ssIDXB3haUmYMpevnnNj/wRRgpcomi7Eo3Fk+Eb7qLk4
qUgQvDMliaxM0Bf2XjRWZBUidQC1DUcuprnRth4XnnIR
lx1pUPq/re/8nQ95pVRS
/sYHWwtOraZlrbECLqhJ40GQLeB5oFJ0ML1MiVfD31xC
jf66kOgI8GAkW0vdgMMPHSr12jjIbyDA+E3+rs1H8g8O
gVhMj7iB4PE+9pnwSi
/aEtwPXzuh6Stbt2t1cuH0ZqZMcyo0tvDLwQit4Qc+06
bBK5CRyu/r6raQbIIS0xddiopvnSSMpihiExample=/
 Amazon-Redshift
```

Node IP Addresses:

| Node | Public IP | Private IP |
|---|---|---|
| Leader | 192.0.2.0 | 198.51.100.0 |
| Compute-0 | 203.0.113.0 | 10.24.34.0 |
| Compute-1 | 198.51.100.0 | 192.0.2.0 |

   En el paso 3, usará las direcciones IP para configurar el alojamiento de forma que acepte la conexión desde Amazon Redshift. Según el tipo de host al que se conecta y si se encuentra en una Virtual Private Cloud (VPC, Nube virtual privada), usará las direcciones IP públicas o las privadas.

Para recuperar la clave pública del clúster y las direcciones IP del nodo del clúster para su clúster mediante la CLI de Amazon Redshift, ejecute el comando describe-clusters.

Por ejemplo:

```
aws redshift describe-clusters --cluster-identifier <cluster-identifier>
```

La respuesta incluirá el valor ClusterPublicKey y la lista de direcciones IP públicas y privadas, similar a la siguiente:

```
{
    "Clusters": [
        {
            "VpcSecurityGroups": [],
            "ClusterStatus": "available",
            "ClusterNodes": [
                {
                    "PrivateIPAddress": "10.nnn.nnn.nnn",
                    "NodeRole": "LEADER",
                    "PublicIPAddress": "10.nnn.nnn.nnn"
                },
                {
                    "PrivateIPAddress": "10.nnn.nnn.nnn",
                    "NodeRole": "COMPUTE-0",
                    "PublicIPAddress": "10.nnn.nnn.nnn"
                },
                {
                    "PrivateIPAddress": "10.nnn.nnn.nnn",
                    "NodeRole": "COMPUTE-1",
                    "PublicIPAddress": "10.nnn.nnn.nnn"
                }
            ],
            "AutomatedSnapshotRetentionPeriod": 1,
            "PreferredMaintenanceWindow": "wed:05:30-wed:06:00",
            "AvailabilityZone": "us-east-1a",
            "NodeType": "dc2.large",
            "ClusterPublicKey": "ssh-rsa AAAABexamplepublickey...Y3TAl Amazon-
Redshift",
             ...
             ...
}
```

Para recuperar la clave pública del clúster y las direcciones IP del nodo del clúster para su clúster mediante la API de Amazon Redshift, use la acción DescribeClusters. Para obtener más información, consulte describe-clusters en la Guía de la CLI de Amazon Redshift o DescribeClusters en la Guía de la API de Amazon Redshift.

## Paso 2: Agregar la clave pública del clúster de Amazon Redshift al archivo de claves autorizadas del alojamiento

Debe agregar la clave pública del clúster al archivo de claves autorizadas de cada alojamiento para que este último reconozca a Amazon Redshift y acepte la conexión SSH.

Para agregar la clave pública del clúster de Amazon Redshift al archivo de claves autorizadas del alojamiento

1. Obtenga acceso al host con una conexión Secure Shell (SSH, Shell seguro).

   Para obtener información acerca de cómo conectarse a una instancia con SSH, consulte Conexión con la instancia en la Guía del usuario de Amazon EC2.
2. Copie la clave pública de Amazon Redshift desde la consola o desde el texto de respuesta de la CLI.
3. Copie y pegue el contenido de la clave pública en el archivo `/home/<ssh_username>/.ssh/authorized_keys` que se encuentra en el host remoto. `<ssh_username>` debe coincidir con el valor del campo "username" del archivo de manifiesto. Incluya la cadena completa, incluso el prefijo "`ssh-rsa`" y el sufijo "`Amazon-Redshift`". Por ejemplo:

   ```
   ssh-rsa AAAACTP3isxgGzVWoIWpbVvRCOzYdVifMrh… uA70BnMHCaMiRdmvsDOedZDOedZ Amazon-
   Redshift
   ```

## Paso 3: Configurar el alojamiento para que acepte todas las direcciones IP del clúster de Amazon Redshift

Si está trabajando con una instancia de Amazon EC2 o con un clúster de Amazon EMR, agregue reglas de entrada al grupo de seguridad del alojamiento para permitir tráfico desde cada nodo del clúster de Amazon Redshift. En Type (Tipo), seleccione SSH with TCP protocol on Port 22 (SSH con protocolo TCP en el puerto 22). En Source (Origen), ingrese las direcciones IP de los nodos del clúster de Amazon Redshift que recuperó en Paso 1: Recuperar la clave pública del clúster y las direcciones IP del nodo del clúster. Para obtener más información acerca de cómo se agregan las reglas a un grupo de seguridad de Amazon EC2, consulte Autorización del tráfico de entrada para sus instancias en la Guía del usuario de Amazon EC2.

Use las direcciones IP privadas en los casos que se muestran a continuación:

- Tiene un clúster de Amazon Redshift que no se encuentra en una Virtual Private Cloud (VPC) y una instancia de Amazon EC2-Classic, y ambas se encuentran en la misma región de AWS.
- Tiene un clúster de Amazon Redshift que se encuentra en una VPC y una instancia de Amazon EC2-VPC, y ambas se encuentran en la misma región de AWS y en la misma VPC.

De lo contrario, use las direcciones IP públicas.

Para obtener más información acerca del uso de Amazon Redshift en una VPC, consulte Administración de clústeres en nube privada virtual (VPC) en la Guía de administración de Amazon Redshift.

## Paso 4: Obtener la clave pública para el host

De manera opcional, puede proporcionar la clave pública del alojamiento en el archivo de manifiesto de manera que Amazon Redshift pueda identificar al alojamiento. El comando COPY no requiere la clave pública del host; no obstante, por motivos de seguridad, le recomendamos encarecidamente que use una clave pública para ayudar a evitar ataques de tipo "man-in-the-middle".

Puede encontrar la clave pública del host en la siguiente ubicación, donde `<ssh_host_rsa_key_name>` es el nombre único de la clave pública del host:

```
:  /etc/ssh/<ssh_host_rsa_key_name>.pub
```

> **Note**
>
> Amazon Redshift solo es compatible con claves RSA. No es compatible con claves DSA.

Al crear el archivo de manifiesto en el paso 5, copiará el texto de la clave pública en el campo "Public Key", en la entrada del archivo de manifiesto.

## Paso 5: Crear un archivo de manifiesto

El comando COPY puede conectarse con distintos hosts mediante Secure Shell (SSH, Shell seguro) y puede crear distintas conexiones SSH a cada host. COPY ejecuta un comando a través de cada conexión al host y, luego, carga la salida en la tabla desde los comandos en paralelo. El archivo de manifiesto es un archivo de texto con formato JSON que Amazon Redshift usa para conectarse al alojamiento. En el archivo de manifiesto, se especifican los puntos de conexión del alojamiento SSH y los comandos que se ejecutan en los alojamientos para devolver los datos a Amazon Redshift. De forma opcional, puede incluir la clave pública del host, el nombre de usuario de inicio de sesión y una marca obligatoria para cada entrada.

Cree el archivo de manifiesto en su equipo local. En un paso posterior, cargará el archivo en Amazon S3.

El archivo de manifiesto se encuentra en el siguiente formato:

```
{
   "entries": [
     {"endpoint":"<ssh_endpoint_or_IP>",
       "command": "<remote_command>",
       "mandatory":true,
       "publickey": "<public_key>",
       "username": "<host_user_name>"},
     {"endpoint":"<ssh_endpoint_or_IP>",
       "command": "<remote_command>",
       "mandatory":true,
       "publickey": "<public_key>",
       "username": "host_user_name"}
    ]
}
```

El archivo de manifiesto tiene una construcción "entries" para cada conexión SSH. Cada entrada representa una conexión SSH única. Puede tener distintas conexiones a un único host o distintas conexiones a distintos hosts. Se requieren comillas dobles, como se muestra, para los valores y los nombres de campo. El único valor que no necesita comillas dobles es el valor booleano **`true`** o **`false`** del campo obligatorio.

A continuación se describen los campos del archivo de manifiesto.

punto de conexión

La dirección URL o dirección IP del host. Por ejemplo, "`ec2-111-222-333.compute-1.amazonaws.com`" o "`22.33.44.56`".

comando

El comando que el alojamiento ejecutará para generar la salida binaria o de texto (en formato gzip, lzop o bzip2). El comando puede ser cualquiera que el usuario "host_user_name" (nombre_de_usuario_del_host) tenga permiso para ejecutar. El comando puede ser tan sencillo como imprimir un archivo, o podría consultar una base de datos o lanzar un script. La salida (archivo de texto o archivo binario gzip, lzop o bzip2) debe estar en un formato que el comando COPY de Amazon Redshift pueda capturar. Para obtener más información, consulte Preparación de los datos de entrada.

publickey

(Opcional) La clave pública del host. Si se proporciona la clave pública, Amazon Redshift la usará para identificar el alojamiento. Si no se proporciona la clave pública, Amazon Redshift no

intentará identificar el alojamiento. Por ejemplo, si la clave pública del host remoto es: `ssh-rsa AbcCbaxxx…xxxDHKJ root@amazon.com`, escriba el siguiente texto en el campo de clave pública: `AbcCbaxxx…xxxDHKJ`.

mandatory

(Opcional) Indica si el comando COPY debe fallar en caso de que la conexión falle. El valor predeterminado es `false`. Si Amazon Redshift no logra al menos una conexión de manera correcta, se produce un error en el comando COPY.

nombre de usuario

(Opcional) Se trata del nombre de usuario que se utilizará para iniciar sesión en el sistema de alojamiento y ejecutar el comando remoto. El nombre de inicio de sesión de usuario debe ser el mismo que el del inicio de sesión que se utilizó para agregar la clave pública al archivo de claves autorizadas del host en el paso 2. El nombre de usuario predeterminado es "redshift".

En el siguiente ejemplo, se observa un manifiesto completado para abrir cuatro conexiones en el mismo alojamiento y ejecutar un comando diferente a través de cada conexión:

```
{
  "entries": [
       {"endpoint":"ec2-184-72-204-112.compute-1.amazonaws.com",
          "command": "cat loaddata1.txt",
          "mandatory":true,
          "publickey": "ec2publickeyportionoftheec2keypair",
          "username": "ec2-user"},
       {"endpoint":"ec2-184-72-204-112.compute-1.amazonaws.com",
          "command": "cat loaddata2.txt",
          "mandatory":true,
          "publickey": "ec2publickeyportionoftheec2keypair",
          "username": "ec2-user"},
       {"endpoint":"ec2-184-72-204-112.compute-1.amazonaws.com",
          "command": "cat loaddata3.txt",
          "mandatory":true,
          "publickey": "ec2publickeyportionoftheec2keypair",
          "username": "ec2-user"},
       {"endpoint":"ec2-184-72-204-112.compute-1.amazonaws.com",
          "command": "cat loaddata4.txt",
          "mandatory":true,
          "publickey": "ec2publickeyportionoftheec2keypair",
          "username": "ec2-user"}
```

```
    ]
}
```

### Paso 6: Cargar el archivo de manifiesto en un bucket de Amazon S3

Cargue el archivo de manifiesto en un bucket de Amazon S3. Si el bucket de Amazon S3 no se encuentra en la misma región de AWS que el clúster de Amazon Redshift, debe usar la opción REGION para especificar la región de AWS en la que se encuentra el manifiesto. Para obtener información acerca de cómo crear un bucket de Amazon S3 y cómo cargar un archivo, consulte Guía del usuario de Amazon Simple Storage Service.

### Paso 7: Ejecutar el comando COPY para cargar los datos

Ejecute un comando COPY para conectarse al alojamiento y cargar los datos en una tabla de Amazon Redshift. En el comando COPY, especifique la ruta de objeto de Amazon S3 de forma explícita para el archivo de manifiesto e incluya la opción para el SSH. Por ejemplo,

```
COPY sales
FROM 's3://amzn-s3-demo-bucket/ssh_manifest'
IAM_ROLE 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
DELIMITER '|'
SSH;
```

> **Note**
>
> Si utiliza la compresión automática, el comando COPY realizará dos lecturas de datos, lo que significa que ejecuta el comando remoto dos veces. La primera lectura se realiza para proporcionar una muestra para el análisis de compresión, la segunda lectura es la que carga los datos. Si la ejecución del comando remoto dos veces puede ocasionar un problema debido a efectos secundarios potenciales, debe desactivar la compresión automática. Para desactivar la compresión automática, ejecute el comando COPY con la opción COMPUPDATE establecida en OFF. Para obtener más información, consulte Carga de tablas con compresión automática.

## Carga de datos desde una tabla de Amazon DynamoDB

Puede usar el comando COPY para cargar una tabla con datos de una tabla específica de Amazon DynamoDB.

> **Important**
>
> La tabla de Amazon DynamoDB que proporciona los datos debe crearse en la misma región de AWS que el clúster, a menos que use la opción REGION para especificar la región de AWS en la que se encuentra la tabla de Amazon DynamoDB.

El comando COPY utiliza la arquitectura de procesamiento masivo en paralelo (MPP) de Amazon Redshift para leer y cargar datos en paralelo desde una tabla de Amazon DynamoDB. Para aprovechar al máximo los beneficios del procesamiento en paralelo, puede establecer estilos de distribución en las tablas de Amazon Redshift. Para obtener más información, consulte Distribución de datos para la optimización de consultas.

> **Important**
>
> Cuando el comando COPY lee los datos desde la tabla de Amazon DynamoDB, la transferencia de datos resultante es parte del rendimiento aprovisionado de esa tabla.

Para evitar el consumo de cantidades excesivas de rendimiento de lectura aprovisionado, le recomendamos que no cargue datos de las tablas de Amazon DynamoDB que se encuentren en entornos de producción. Si efectivamente carga datos desde tablas de producción, le recomendamos que establezca la opción READRATIO en un valor mucho más bajo que el porcentaje promedio de rendimiento provisionado no usado. Una configuración READRATIO baja ayudará a minimizar los problemas de limitación controlada. Para usar el rendimiento aprovisionado completo de una tabla de Amazon DynamoDB, establezca el valor de READRATIO en 100.

El comando COPY establece una concordancia entre los nombres de atributo de los elementos recuperados de la tabla de DynamoDB y los nombres de columna de una tabla de Amazon Redshift existente mediante las siguientes reglas:

- Las columnas de la tabla de Amazon Redshift establecen una concordancia, sin distinción entre mayúsculas y minúsculas, con los atributos de los elementos de Amazon DynamoDB. Si un elemento en la tabla de DynamoDB tiene varios atributos que difieren solo en las mayúsculas y minúsculas, como Price y PRICE, el comando COPY provocará un error.
- Las columnas de la tabla de Amazon Redshift que no concuerdan con ningún atributo de la tabla de Amazon DynamoDB se cargan como NULL o empty, según el valor especificado en la opción EMPTYASNULL del comando COPY.

- Los atributos de Amazon DynamoDB que no concuerdan con ninguna columna de la tabla de Amazon Redshift se descartan. Los atributos se leen antes de hacerlos coincidir, por lo que incluso los atributos descartados consumen parte del rendimiento provisionado de esa tabla.
- Solo se admiten los atributos de Amazon DynamoDB con tipos de datos STRING y NUMBER escalares. No se admiten los tipos de datos BINARY y SET de Amazon DynamoDB. Si un comando COPY prueba cargar un atributo con un tipo de dato incompatible, el comando provocará un error. Si el atributo no concuerda con una columna de la tabla de Amazon Redshift, COPY no intenta cargarlo y no genera un error.

El comando COPY usa la siguiente sintaxis para cargar datos desde una tabla de Amazon DynamoDB:

```
COPY <redshift_tablename> FROM 'dynamodb://<dynamodb_table_name>'
authorization
readratio '<integer>';
```

Los valores de authorization (autorización) son las credenciales de AWS necesarias para obtener acceso a la tabla de Amazon DynamoDB. Si estas credenciales corresponden a un usuario, este último debe tener permiso para el uso de SCAN y DESCRIBE en la tabla de Amazon DynamoDB que se está cargando.

Los valores de authorization (autorización) proporcionan la autorización de AWS que el clúster necesita para obtener acceso a la tabla de Amazon DynamoDB. El permiso debe incluir el uso de SCAN y DESCRIBE en la tabla de Amazon DynamoDB que se está cargando. Para obtener más información acerca de los permisos requeridos, consulte Permisos de IAM para COPY, UNLOAD y CREATE LIBRARY. El método de autenticación preferido es especificar el parámetro IAM_ROLE y proporcionar el Amazon Resource Name (ARN, Nombre de recurso de Amazon) de un rol de IAM con los permisos necesarios. Para obtener más información, consulte Control de acceso con base en roles.

Para realizar la autenticación con el parámetro IAM_ROLE, reemplace *<aws-account-id>* y *<role-name>*, tal y como se observa en la siguiente sintaxis.

```
IAM_ROLE 'arn:aws:iam::<aws-account-id>:role/<role-name>'
```

En el siguiente ejemplo, se muestra la autenticación con un rol de IAM.

```
COPY favoritemovies
```

```
FROM 'dynamodb://ProductCatalog'
IAM_ROLE 'arn:aws:iam::0123456789012:role/MyRedshiftRole';
```

Para obtener más información acerca de otras opciones de autorización, consulte Parámetros de autorización.

Si desea validar los datos sin cargar de hecho la tabla, use la opción NOLOAD con el comando COPY.

En el siguiente ejemplo, la tabla FAVORITEMOVIES se carga con datos de la tabla de DynamoDB my-favorite-movies-table. La actividad de lectura puede consumir hasta el 50% del rendimiento provisionado.

```
COPY favoritemovies FROM 'dynamodb://my-favorite-movies-table'
IAM_ROLE 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
READRATIO 50;
```

Para maximizar el rendimiento, el comando COPY carga los datos desde una tabla de Amazon DynamoDB en paralelo en todos los nodos informáticos del clúster.

### Rendimiento aprovisionado con compresión automática

De forma predeterminada, el comando COPY aplica una compresión automática siempre que especifique una tabla destino vacía sin cifrado de compresión. En un principio, el análisis de compresión automático realiza una muestra de una gran cantidad de filas de la tabla de Amazon DynamoDB. El tamaño de la muestra se basa en el valor del parámetro COMPROWS. El valor predeterminado es 100 000 filas por sector.

Una vez terminado el muestreo, las filas de muestra se descartan y se carga la tabla completa. Como resultado, muchas filas se leen dos veces. Para obtener más información acerca de cómo funciona la compresión automática, consulte Carga de tablas con compresión automática.

> **Important**
>
> Cuando el comando COPY lee datos de la tabla de Amazon DynamoDB, incluidas las filas usadas para el muestreo, la transferencia de datos resultante es parte del rendimiento aprovisionado de esa tabla.

### Carga de datos multibyte desde Amazon DynamoDB

Si los datos incluyen caracteres multibyte no ASCII (como caracteres cirílicos o chinos) debe cargar los datos en columnas VARCHAR. El tipo de datos VARCHAR admite caracteres UTF-8 de cuatro bytes, pero el tipo de datos CHAR solo acepta caracteres ASCII de un solo byte. No puede cargar caracteres de cinco bytes o más en tablas de Amazon Redshift. Para obtener más información acerca de los comandos CHAR y VARCHAR, consulte Tipos de datos.

## Comprobación de carga correcta de datos

Luego de que se complete la operación de carga, consulte la tabla de sistema STL_LOAD_COMMITS para controlar que se hayan cargado los archivos esperados. Ejecute el comando COPY y la verificación de carga en la misma transacción, de manera que si se produce un problema con la carga, se pueda revertir toda la transacción.

La siguiente consulta devuelve entradas para la carga de las tablas en la base de datos TICKIT:

```
SELECT query, trim(filename) AS filename, curtime, status
FROM stl_load_commits
WHERE filename like '%tickit%' order by query;


 query |         filename          |          curtime           | status
-------+---------------------------+----------------------------+--------
 22475 | tickit/allusers_pipe.txt  | 2013-02-08 20:58:23.274186 |      1
 22478 | tickit/venue_pipe.txt     | 2013-02-08 20:58:25.070604 |      1
 22480 | tickit/category_pipe.txt  | 2013-02-08 20:58:27.333472 |      1
 22482 | tickit/date2008_pipe.txt  | 2013-02-08 20:58:28.608305 |      1
 22485 | tickit/allevents_pipe.txt | 2013-02-08 20:58:29.99489  |      1
 22487 | tickit/listings_pipe.txt  | 2013-02-08 20:58:37.632939 |      1
 22489 | tickit/sales_tab.txt      | 2013-02-08 20:58:37.632939 |      1
(6 rows)
```

## Validación de datos de entrada

Para validar los datos de los archivos de entrada de Amazon S3 o de la tabla de Amazon DynamoDB antes de que los datos se carguen, use la opción NOLOAD con el comando COPY. Use NOLOAD con los mismos comandos COPY y con las mismas opciones que usaría para cargar los datos. NOLOAD comprueba la integridad de todos los datos sin cargarlos en la base de datos. La opción NOLOAD muestra los errores que se producen si intenta cargar los datos.

Por ejemplo, si especificó una ruta de Amazon S3 incorrecta para el archivo de entrada, Amazon Redshift mostrará el siguiente error.

```
ERROR:  No such file or directory
DETAIL:
-----------------------------------------------
Amazon Redshift error:  The specified key does not exist
code:      2
context:   S3 key being read :
location:  step_scan.cpp:1883
process:   xenmaster [pid=22199]
-----------------------------------------------
```

Para solucionar problemas con los mensajes de error, consulte Referencia de error de carga.

Para ver un ejemplo de uso de la opción NOLOAD, consulte Comando COPY con la opción NOLOAD.

## Carga de tablas con compresión automática

Puede aplicar codificaciones de compresión manualmente a las columnas de las tablas, en función de su propia evaluación de los datos. También puede utilizar el comando COPY con COMPUPDATE establecido en ON para analizar y aplicar la compresión de forma automática en función de los datos de muestra.

Puede usar la compresión automática al crear y cargar una tabla completamente nueva. El comando COPY realiza un análisis de compresión. También puede realizar un análisis de compresión sin cargar datos ni cambiar la compresión de una tabla ejecutando el comando ANALYZE COMPRESSION en una tabla que ya contenga datos. Por ejemplo, puede ejecutar ANALYZE COMPRESSION si desea analizar la compresión de una tabla para utilizarla más adelante y presentar las instrucciones del lenguaje de definición de datos (DDL) existente.

La compresión automática equilibra el rendimiento general al elegir las codificaciones de compresión. Los exámenes de rango restringido podrían tener un rendimiento deficiente cuando las columnas de clave con ordenación se comprimen mucho más que otras columnas en la misma consulta. Como resultado, la compresión automática omitirá la fase de análisis de datos en las columnas con la clave de ordenación y mantendrá los tipos de codificación definidos por el usuario.

La compresión automática elige la codificación RAW si no se ha definido explícitamente ningún tipo de codificación. ANALYZE COMPRESSION hace lo mismo. Para obtener un rendimiento óptimo de las consultas, considere la posibilidad de utilizar RAW con las claves de ordenación.

## Cómo funciona la compresión automática

Cuando el parámetro COMPUDATE está establecido en ON, el comando COPY aplica la compresión automática siempre que se ejecute con una tabla de destino y todas las columnas de la tabla tengan codificación RAW o no tengan codificación.

Para aplicar la compresión automática en una tabla vacía, independientemente de sus codificaciones de compresión actuales, ejecute el comando COPY con la opción COMPUPDATE establecida en ON. Para desactivar la compresión automática, ejecute el comando COPY con la opción COMPUPDATE establecida en OFF.

No puede aplicar la compresión automática a una tabla que ya tiene datos.

> **Note**
>
> El análisis de compresión automático requiere suficientes filas en los datos cargados (al menos 100 000 filas por sector) con objeto de generar una muestra significativa.

La compresión automática realiza estas operaciones en segundo plano como parte de la transacción de carga:

1. Una muestra inicial de filas se carga desde el archivo de entrada. El tamaño de la muestra se basa en el valor del parámetro COMPROWS. El valor predeterminado es 100,000.
2. Las opciones de compresión se seleccionan para cada columna.
3. Las filas de muestra se eliminan de la tabla.
4. La tabla se recrea con las codificaciones de compresión seleccionadas.
5. El archivo de entrada completo se carga y se comprime con las codificaciones nuevas.

Una vez ejecutado el comando COPY, la tabla se encuentra completamente cargada, comprimida y lista para su uso. Si carga más datos con posterioridad, las filas agregadas se comprimen según la codificación existente.

Si solo desea realizar un análisis de compresión, ejecute ANALYZE COMPRESSION, lo cual es más eficiente que ejecutar COPY por completo. Luego, puede evaluar los resultados para decidir si usar la compresión automática o recrear la tabla manualmente.

La compresión automática es compatible solo para el comando COPY. De manera alternativa, puede aplicar la codificación de compresión manualmente al crear la tabla. Para obtener información acerca de la codificación de compresión manual, consulte Compresión de columnas para reducir el tamaño de los datos almacenados.

## Ejemplo de compresión automática

En este ejemplo, se asume que la base de datos TICKIT tiene una copia de la tabla LISTING denominada BIGLIST y que usted desea aplicar la compresión automática en esta tabla una vez cargada con, aproximadamente, 3 millones de filas.

Pasos para cargar y comprimir una tabla automáticamente

1. Asegúrese de que la tabla esté vacía. Puede aplicar la compresión automática solamente a una tabla vacía:

   ```
   TRUNCATE biglist;
   ```

2. Cargue la tabla con un comando COPY único. Aunque la tabla esté vacía, cierta codificación podría haberse especificado con anterioridad. Para facilitar que Amazon Redshift realice un análisis de compresión, establezca el parámetro COMPUPDATE en ON.

   ```
   COPY biglist FROM 's3://amzn-s3-demo-bucket/biglist.txt'
   IAM_ROLE 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
   DELIMITER '|' COMPUPDATE ON;
   ```

   Como no se especifica una opción COMPROWS, se usa el tamaño de la muestra predeterminado y recomendado, que es 100 000 filas por sector.

3. Observe el esquema nuevo para la tabla BIGLIST con objeto de revisar los esquemas de codificación seleccionados de manera automática.

   ```
   SELECT "column", type, encoding
   from pg_table_def where tablename = 'biglist';


        Column     |            Type             | Encoding
   ----------------+-----------------------------+----------
    listid         | integer                     | az64
    sellerid       | integer                     | az64
    eventid        | integer                     | az64
   ```

```
 dateid         | smallint                    | none
 numtickets     | smallint                    | az64
 priceperticket | numeric(8,2)                | az64
 totalprice     | numeric(8,2)                | az64
 listtime       | timestamp without time zone | az64
```

4. Compruebe que se haya cargado la cantidad de filas esperada:

```
select count(*) from biglist;

count
---------
3079952
(1 row)
```

Cuando, posteriormente, se agreguen filas a esta tabla mediante instrucciones COPY o INSERT, se aplican las mismas codificaciones de compresión.

## Optimización del almacenamiento para tablas angostas

Si tiene una tabla con muy pocas columnas, pero con una gran cantidad de filas, las tres columnas de identidad de metadatos ocultas (INSERT_XID, DELETE_XID, ROW_ID) consumirán una cantidad desproporcionada del espacio en disco para la tabla.

Para optimizar la compresión de las columnas ocultas, cargue la tabla en una transacción COPY única siempre que sea posible. Si carga la tabla con distintos comandos COPY por separado, la columna INSERT_XID no se comprimirá bien. Si usa distintos comandos COPY, debe realizar una operación de limpieza, pero no mejorará la compresión de INSERT_XID.

## Carga de valores de columna predeterminados

De manera opcional, puede definir una lista de columnas en su comando COPY. Si una columna de la tabla se omite en la lista de columnas, COPY la copiará o con el valor proporcionado por la opción DEFAULT especificada en el comando CREATE TABLE o con NULL, si la opción DEFAULT no se especificó.

Si COPY prueba asignar NULL a una columna que está definida como NOT NULL, el comando COPY falla. Para obtener información acerca de asignar la opción DEFAULT, consulte CREATE TABLE.

Cuando se realiza la carga desde los archivos de datos en Amazon S3, las columnas de la lista de columnas deben encontrarse en el mismo orden en que se encuentran los campos en el archivo de datos. Si un campo en el archivo de datos no tiene una columna correspondiente en la lista de columnas, el comando COPY provocará un error.

Cuando se realiza la carga desde una tabla de Amazon DynamoDB, el orden no importa. Los campos en los atributos de Amazon DynamoDB que no concuerdan con ninguna columna de la tabla de Amazon Redshift se descartan.

Al usar el comando COPY para cargar valores DEFAULT en una tabla se aplican las siguientes restricciones:

- Si se incluye una columna IDENTITY en la lista de columnas, también se debe especificar la opción EXPLICIT_IDS en el comando COPY, o este último provocará un error. De manera similar, si se omite una columna IDENTITY de la lista de columnas y se especifica la opción EXPLICIT_IDS, la operación COPY provocará un error.
- Como la expresión DEFAULT evaluada para una columna dada es la misma para todas las filas cargadas, una expresión DEFAULT que usa una función RANDOM() asignará el mismo valor a todas las filas.
- Las expresiones DEFAULT que tienen CURRENT_DATE o SYSDATE se establecen según la marca temporal de la transacción actual.

Para obtener un ejemplo, consulte "Load data from a file with default values" en Ejemplos de COPY.

## Solución de problemas en cargas de datos

Al cargar datos en tablas de Amazon Redshift, es posible que encuentre errores de Amazon S3, datos de entrada no válidos y errores del comando COPY. En las siguientes secciones, se ofrece información acerca de cómo identificar y resolver errores en la carga de datos.

Temas

- Solución de los errores de la integración de eventos de S3 y COPY JOB
- Errores S3ServiceException
- Tablas de sistema para la solución de problemas de cargas de datos
- Errores de carga de caracteres multibyte
- Referencia de error de carga

## Solución de los errores de la integración de eventos de S3 y COPY JOB

Utilice la siguiente información para solucionar problemas habituales de la integración de eventos de Amazon S3 y COPY JOB con Amazon Redshift.

Error en la creación de la integración de eventos de S3

Si se ha producido un error en la creación de la integración de eventos de S3, el estado de la integración será `Inactive`. Asegúrese de que lo siguiente sea correcto para el almacenamiento de datos de Amazon Redshift.

- Ha añadido la entidad principal autorizada y el origen de la integración correctos para su espacio de nombres de destino en Amazon Redshift. Consulte Requisitos previos para crear una integración de eventos de S3.
- Ha añadido la política basada en recursos correcta al bucket de Amazon S3 de origen. Consulte Requisitos previos para crear una integración de eventos de S3.

Los datos de Amazon S3 no aparecen en la base de datos de destino

Si no aparecen los datos de un COPY JOB, compruebe lo siguiente.

- Consulte SYS_COPY_JOB_DETAIL para ver si el archivo de Amazon S3 se ha cargado, si está pendiente de ingesta o si hay algún error. Para obtener más información, consulte SYS_COPY_JOB_DETAIL.
- Consulte STL_ERROR o SYS_COPY_JOB_INFO si el archivo de Amazon S3 no está ahí o si se produce un tiempo de espera inesperado. Busque errores de credenciales o cualquier cosa que sugiera que la integración está inactiva. Para obtener más información, consulte STL_ERROR y SYS_COPY_JOB_INFO.

## Errores S3ServiceException

Los errores s3ServiceException más comunes son causados por una cadena de credenciales incorrecta o sin el formato adecuado, la presencia del clúster y el bucket en regiones de AWS diferentes y permisos de Amazon S3 insuficientes.

En esta sección, se proporciona información acerca de cómo solucionar problemas para cada tipo de error.

Cadena de credenciales no válida

Si la cadena de credenciales no tiene el formato adecuado, recibirá el siguiente mensaje de error:

```
ERROR: Invalid credentials. Must be of the format: credentials
'aws_access_key_id=<access-key-id>;aws_secret_access_key=<secret-access-key>
[;token=<temporary-session-token>]'
```

Compruebe que la cadena de credenciales no tenga espacios ni saltos de línea y que se encuentre delimitada entre comillas simples.

ID de clave de acceso no válida

Si el ID de clave de acceso no existe, recibirá el siguiente mensaje de error:

```
[Amazon](500310) Invalid operation: S3ServiceException:The AWS Access Key Id you
 provided does not exist in our records.
```

Este suele ser un error de copiado y pegado. Compruebe que el ID de clave de acceso se haya escrito correctamente. Además, si utiliza claves de sesiones temporales, compruebe que esté definido el valor de `token`.

Clave de acceso secreta no válida

Si su clave de acceso secreta es incorrecta, recibirá el siguiente mensaje de error:

```
[Amazon](500310) Invalid operation: S3ServiceException:The request signature we
 calculated does not match the signature you provided.
Check your key and signing method.,Status 403,Error SignatureDoesNotMatch
```

Este suele ser un error de copiado y pegado. Compruebe que la clave de acceso secreta se haya escrito correctamente y que sea la clave correcta para el ID de clave de acceso.

El bucket está en una región diferente

El bucket de Amazon S3 especificado en el comando COPY debe estar en la misma región de AWS que el clúster. Si su bucket de Amazon S3 y su clúster están en regiones diferentes, se producirá un error similar al siguiente:

```
ERROR: S3ServiceException:The bucket you are attempting to access must be addressed
 using the specified endpoint.
```

Puede crear un bucket de Amazon S3 en una región específica. Para ello, seleccione la región mientras crea el bucket con la consola de administración de Amazon S3 o especifique un punto de conexión mientras crea el bucket con la CLI o la API de Amazon S3. Para obtener más información, consulte Carga de archivos en Amazon S3 para utilizar con COPY.

Para obtener más información acerca de las regiones de Amazon S3, consulte Acceso a un bucket en la Guía del usuario de Amazon Simple Storage Service.

También puede especificar la región utilizando la opción REGION con el comando COPY.

Acceso denegado

Si el usuario no tiene permisos suficientes, recibirá el siguiente mensaje de error:

```
ERROR: S3ServiceException:Access Denied,Status 403,Error AccessDenied
```

Una causa posible es que el usuario identificado por las credenciales no tiene acceso LIST y GET al bucket de Amazon S3. Para otras causas, consulte Solucionar errores de acceso denegado (403 Prohibido) en Amazon S3 en la Guía del usuario de Amazon Simple Storage Service.

Para obtener información sobre cómo administrar el acceso de los usuarios a los buckets, consulte Administración de identidades y accesos en Amazon S3 en la Guía del usuario de Amazon Simple Storage Service.

## Tablas de sistema para la solución de problemas de cargas de datos

Las siguientes tablas del sistema de Amazon Redshift pueden ser útiles para solucionar problemas relacionados con la carga de datos:

- Ejecute la consulta STL_LOAD_ERRORS para encontrar los errores que ocurrieron durante cargas específicas.
- Ejecute la consulta STL_FILE_SCAN para ver los tiempos de carga de archivos específicos o para ver si un archivo específico siquiera se leyó.
- Ejecute la consulta STL_S3CLIENT_ERROR para encontrar detalles sobre los errores que surgieron mientras se transferían datos desde Amazon S3.

Pasos para encontrar y diagnosticar errores de carga

1. Cree una vista o defina una consulta que devuelve detalles sobre errores de carga. En el siguiente ejemplo, se combina la tabla STL_LOAD_ERRORS y la tabla STV_TBL_PERM para hacer coincidir los ID de tabla con los nombres reales de tabla.

```
create view loadview as
(select distinct tbl, trim(name) as table_name, query, starttime,
trim(filename) as input, line_number, colname, err_code,
trim(err_reason) as reason
from stl_load_errors sl, stv_tbl_perm sp
where sl.tbl = sp.id);
```

2. Establezca la opción MAXERRORS en el comando COPY en un valor lo suficientemente grande como para permitirle a COPY que devuelva información útil sobre sus datos. Si COPY encuentra errores, un mensaje de error le indica que consulte la tabla STL_LOAD_ERRORS para obtener detalles.
3. Consulte la vista LOADVIEW para obtener detalles del error. Por ejemplo:

```
select * from loadview where table_name='venue';
```

```
  tbl   | table_name | query |         starttime
--------+------------+-------+----------------------------
 100551 | venue      | 20974 | 2013-01-29 19:05:58.365391

|     input      | line_number | colname | err_code |       reason
+----------------+-------------+---------+----------+--------------------
| venue_pipe.txt |           1 |       0 |     1214 | Delimiter not found
```

4. Corrija el problema en el archivo de entrada o en el script de carga, en función de la información que devuelva la vista. Entre los errores de carga habituales por buscar se incluyen los siguientes:

   - Discrepancia entre los tipos de datos de la tabla y los valores de los campos de datos de entrada.
   - Discrepancia entre la cantidad de columnas de la tabla y la cantidad de campos de datos de entrada.

- Discrepancia en las comillas. Amazon Redshift admite comillas simples y dobles; no obstante, las comillas deben equilibrarse de forma adecuada.
- Formato incorrecto para los datos de fecha/hora en los archivos de entrada.
- Valores fuera de rango en los archivos de entrada (para las columnas numéricas).
- La cantidad de valores distintos para una columna supera la limitación para su codificación de compresión.

## Errores de carga de caracteres multibyte

Las columnas con un tipo de datos CHAR aceptan solamente caracteres UTF-8 de un solo byte, hasta un valor de byte 127 o 7F hexadecimal, que también es el conjunto de caracteres ASCII. Las columnas VARCHAR aceptan caracteres multibyte UTF-8 de hasta un máximo de cuatro bytes. Para obtener más información, consulte Tipos de caracteres.

Si una línea de los datos cargados contiene un carácter no válido para el tipo de datos de la columna, COPY devuelve un error y registra una fila en la tabla de registro del sistema STL_LOAD_ERRORS con el número de error 1220. El campo ERR_REASON incluye la secuencia de bytes, en hexadecimal, para el carácter no válido.

Una alternativa a corregir caracteres no válidos en los datos de carga es sustituir los caracteres no válidos durante el proceso de carga. Para reemplazar los caracteres UTF-8 no válidos, especifique la opción ACCEPTINVCHARS del comando COPY. Si se ha configurado la opción ACCEPTINVCHARS, el carácter que especifique sustituye al punto de código. Si la opción ACCEPTINVCHARS no está configurada, Amazon Redshift acepta los caracteres como UTF-8 válido. Para obtener más información, consulte ACCEPTINVCHARS.

La siguiente lista de puntos de código es UTF-8 válida, las operaciones COPY no devuelven un error si la opción ACCEPTINVCHARS no está configurada. No obstante, estos puntos de código son caracteres no válidos. Puede utilizar la opción ACCEPTINVCHARS para reemplazar un punto de código por un carácter que especifique. Estos puntos de código incluyen el rango de valores de `0xFDD0` a `0xFDEF` y valores hasta `0x10FFFF`, terminando con `FFFE` o `FFFF`:

- `0xFFFE`, `0x1FFFE`, `0x2FFFE`, ..., `0xFFFFE`, `0x10FFFE`
- `0xFFFF`, `0x1FFFF`, `0x2FFFF`, ..., `0xFFFFF`, `0x10FFFF`

En el siguiente ejemplo, se muestra la razón del error cuando COPY intenta cargar el carácter UTF-8 `e0 a1 c7a4` en una columna CHAR.

```
Multibyte character not supported for CHAR
(Hint: Try using  VARCHAR). Invalid char: e0 a1 c7a4
```

Si el error se relaciona con un tipo de datos VARCHAR, la razón del error incluye un código de error y la secuencia hexadecimal UTF-8 no válida. En el siguiente ejemplo, se muestra la razón del error cuando COPY prueba cargar el carácter UTF-8 a4 en un campo VARCHAR.

```
String contains invalid or unsupported UTF-8 codepoints.
Bad UTF-8 hex sequence: a4 (error 3)
```

En la siguiente tabla, se presentan las descripciones y las soluciones a problemas sugeridas para los errores de carga VARCHAR. Si se produce uno de estos errores, reemplace el carácter por una secuencia de código UTF-8 válida o elimínelo.

| Código de error | Descripción |
|---|---|
| 1 | La secuencia de bytes UTF-8 supera el máximo de cuatro bytes compatible con VARCHAR. |
| 2 | La secuencia de bytes UTF-8 está incompleta. COPY no encontró la cantidad esperada de bytes de continuación para un carácter multibyte antes del final de la cadena. |
| 3 | El carácter de un byte UTF-8 se encuentra fuera de rango. El byte de comienzo no debe ser 254, 255 ni cualquier carácter entre 128 y 191 (inclusive). |
| 4 | El valor del byte a la derecha en la secuencia de bytes se encuentra fuera de rango. El byte de continuación debe encontrarse entre 128 y 191 (inclusive). |
| 5 | El carácter UTF-8 se reserva como sustituto. Los valores de punto sustitutos (de U +D800 a U+DFFF) no son válidos. |
| 8 | La secuencia de bytes supera el valor de punto UTF-8 máximo. |
| 9 | La secuencia de bytes UTF-8 no tiene un valor de punto correspondiente. |

## Referencia de error de carga

Si se producen errores mientras se cargan datos desde un archivo, ejecute una consulta en la tabla STL_LOAD_ERRORS para identificar el error y determinar una explicación posible. La siguiente tabla presenta todos los códigos de error que podrían producirse durante cargas de datos:

Códigos de error de carga

| Código de error | Descripción |
|---|---|
| 1200 | Error de análisis desconocido. Póngase en contacto con el soporte. |
| 1201 | No se encontró el delimitador de campo en el archivo de entrada. |
| 1202 | Los datos de entrada tuvieron más columnas que las definidas en el DDL. |
| 1203 | Los datos de entrada tuvieron menos columnas que las definidas en el DDL. |
| 1204 | Los datos de entrada superaron el rango aceptable para el tipo de dato. |
| 1205 | El formato de la fecha no es válido. Para conocer los formatos válidos, consulte Cadenas TIMEFORMAT y DATEFORMAT. |
| 1206 | El formato de la marca temporal no es válido. Para conocer los formatos válidos, consulte Cadenas TIMEFORMAT y DATEFORMAT. |
| 1207 | Los datos presentan un valor fuera del rango esperado de 0 a 9. |
| 1208 | Error de formato de tipo de datos FLOAT. |
| 1209 | Error de formato de tipo de datos DECIMAL. |
| 1210 | Error de formato de tipo de datos BOOLEAN. |
| 1211 | La línea de entrada no presenta datos. |
| 1212 | No se encontró el archivo de carga. |
| 1213 | Un campo especificado como NOT NULL no presenta datos. |

| Código de error | Descripción |
|---|---|
| 1214 | No se encontró el delimitador. |
| 1215 | Error en el campo CHAR. |
| 1216 | La línea de entrada no es válida. |
| 1217 | El valor de la columna de identidad no es válido. |
| 1218 | Al usar NULL AS '\0', un campo que tiene un carácter de terminación (NUL o UTF-8 0000) presenta más de un byte. |
| 1219 | El hexadecimal UTF-8 tiene un dígito no válido. |
| 1220 | La cadena tiene valores de punto UTF-8 no compatibles o no válidos. |
| 1221 | La codificación del archivo no es la especificada en el comando COPY. |
| 1222 | Error de desbordamiento de valores enteros. |
| 1223 | El tipo de datos no es válido. |
| 1224 | Los datos de entrada no tienen un formato JSON correcto o son de tipo de datos SUPER. |
| 8001 | El parámetro COPY with MANIFEST requiere la ruta completa de un objeto de Amazon S3. |
| 9005 | Se ha especificado una clave final no válida. |

# Creación de una integración de eventos de S3 para copiar automáticamente archivos desde buckets de Amazon S3

> **Note**
>
> La versión preliminar de la copia automática ha finalizado. Por consiguiente, los clústeres de la versión preliminar se eliminarán automáticamente 30 días después de que finalice

el periodo de la versión preliminar. Si tiene pensado seguir utilizando la copia automática, le recomendamos que vuelva a crear sus trabajos de copia automática existentes en otro clúster de Amazon Redshift. No se admite la actualización de un clúster de la versión preliminar a la versión más reciente de Amazon Redshift.

Puede utilizar un trabajo de copia automática para cargar datos en las tablas de Amazon Redshift desde archivos almacenados en Amazon S3. Amazon Redshift detecta cuándo se agregan nuevos archivos de Amazon S3 a la ruta especificada en el comando COPY. A continuación, se ejecuta automáticamente un comando COPY sin tener que crear una canalización de ingesta de datos externa. Amazon Redshift hace un seguimiento de los archivos que se han cargado. Amazon Redshift determina la cantidad de archivos agrupados en lotes por comando COPY. Puede ver los comandos COPY resultantes en las vistas del sistema.

El primer paso para crear un COPY JOB automático es crear una integración de eventos de S3. Cuando aparece un archivo nuevo en el bucket de origen de Amazon S3, Amazon Redshift administra la carga de los archivos en la base de datos mediante el comando COPY.

## Requisitos previos para crear una integración de eventos de S3

Para configurar la integración de eventos de S3, confirme que se cumplen los siguientes requisitos previos.

- El bucket de Amazon S3 debe tener una política de bucket que permita varios permisos de Amazon S3. Por ejemplo, la siguiente política de ejemplo da permisos al bucket de recursos `amzn-s3-demo-bucket` que está alojado en *`us-east-1`*. Tanto el bucket de Amazon S3 como la integración están en la misma Región de AWS.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "Auto-Copy-Policy-01",
            "Effect": "Allow",
            "Principal": {
                "Service": "redshift.amazonaws.com"
                },
            "Action": [
                "s3:GetBucketNotification",
                "s3:PutBucketNotification",
```

```
                "s3:GetBucketLocation"
            ],
            "Resource": "arn:aws:s3:::amzn-s3-demo-bucket",
            "Condition": {
                "StringLike": {
                    "aws:SourceArn": "arn:aws:redshift:us-
east-1:123456789012:integration:*",
                    "aws:SourceAccount": "123456789012"
                }
            }
        }
    ]
}
```

- El clúster aprovisionado de Amazon Redshift de destino o el espacio de nombres de Redshift sin servidor deben tener permiso para el bucket. Confirme que un rol de IAM asociado a su clúster o espacio de nombres sin servidor tiene una política de IAM que concede los permisos adecuados. La política debe permitir tanto `s3:GetObject` para un recurso de bucket como *amzn-s3-demo-bucket* y `s3:ListBucket` para un recurso de bucket y su contenido como, por ejemplo, *amzn-s3-demo-bucket*`/*`.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AutoCopyReadId",
            "Effect": "Allow",
            "Action": [
                "s3:GetObject",
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::amzn-s3-demo-bucket",
                "arn:aws:s3:::amzn-s3-demo-bucket/*"
            ]
        }
    ]
}
```

Añada la política a un rol de IAM que tenga una relación de confianza para el rol de la siguiente manera.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": [
                    "redshift.amazonaws.com"
                ]
            },
            "Action": "sts:AssumeRole"
        }
    ]
}
```

Si el almacenamiento de datos de destino es un clúster aprovisionado, puede asociar un rol de IAM al clúster aprovisionado mediante la consola de Amazon Redshift en la pestaña Permisos del clúster de los detalles del clúster. Para obtener información sobre cómo asociar un rol al clúster aprovisionado, consulte Asociación de roles de IAM a los clústeres en la Guía de administración de Amazon Redshift.

Si el almacenamiento de datos de destino es Redshift sin servidor, puede asociar un rol de IAM al espacio de nombres sin servidor mediante la consola de Redshift sin servidor en la pestaña Seguridad y cifrado de los detalles del espacio de nombres. Para obtener información sobre cómo asociar un rol al espacio de nombres sin servidor, consulte Concesión de permisos a Amazon Redshift Serverless en la Guía de administración de Amazon Redshift.

- Su almacenamiento de datos de Amazon Redshift también debe tener una política de recursos que permita el bucket de Amazon S3. Si utiliza la consola de Amazon Redshift, al crear la integración de eventos de S3, Amazon Redshift ofrece la opción Arréglalo por mí para añadir esta política al almacenamiento de datos de Amazon Redshift. Para actualizar una política de recursos por su cuenta, puede usar el comando put-resource-policy de la AWS CLI. Por ejemplo, para asociar una política de recursos al clúster aprovisionado de Amazon Redshift para una integración de eventos de S3 con un bucket de Amazon S3, ejecute un comando de la AWS CLI similar al siguiente. En el siguiente ejemplo, se muestra una política para un espacio de nombres del clúster aprovisionado en la Región de AWS *us-east-1* para la cuenta de usuario *123456789012*. El bucket se denomina *amzn-s3-demo-bucket*.

```
aws redshift put-resource-policy \
```

```
--policy file://rs-rp.json \
--resource-arn "arn:aws:redshift: us-east-1:123456789012:namespace/cc4ffe56-
ad2c-4fd1-a5a2-f29124a56433"
```

Donde `rs-rp.json` contiene:

```
{
 "Version": "2012-10-17",
 "Statement": [
  {
   "Effect": "Allow",
   "Principal": {
    "Service": "redshift.amazonaws.com"
   },
   "Action": "redshift:AuthorizeInboundIntegration",
   "Resource": "arn:aws:redshift:us-east-1:123456789012:namespace/cc4ffe56-ad2c-4fd1-
a5a2-f29124a56433",
   "Condition": {
    "StringEquals": {
     "aws:SourceArn": "arn:aws:s3:::amzn-s3-demo-bucket"
    }
   }
  },
  {
   "Effect": "Allow",
   "Principal": {
    "AWS": "arn:aws:iam::123456789012:role/myRedshiftRole"
   },
   "Action": "redshift:CreateInboundIntegration",
   "Resource": "arn:aws:redshift:us-east-1:123456789012:namespace/cc4ffe56-ad2c-4fd1-
a5a2-f29124a56433",
   "Condition": {
    "StringEquals": {
     "aws:SourceArn": "arn:aws:s3:::amzn-s3-demo-bucket"
    }
   }
  }
 ]
}
```

Para asociar una política de recursos a su espacio de nombres de Redshift sin servidor para una integración de eventos de S3 con un bucket de Amazon S3, ejecute un comando de la AWS CLI

similar al siguiente. En el siguiente ejemplo, se muestra una política para un espacio de nombres sin servidor en la Región de AWS *us-east-1* para la cuenta de usuario *123456789012*. El bucket se denomina *amzn-s3-demo-bucket*.

```
aws redshift put-resource-policy \
--policy file://rs-rp.json \
--resource-arn "arn:aws:redshift-serverless:us-east-1:123456789012:namespace/
namespace-1"
```

Donde `rs-rp.json` contiene:

```
{
 "Version": "2012-10-17",
 "Statement": [
  {
   "Effect": "Allow",
   "Principal": {
    "Service": "redshift.amazonaws.com"
   },
   "Action": "redshift:AuthorizeInboundIntegration",
   "Resource": "arn:aws:redshift-serverless:us-east-1:123456789012:namespace/
namespace-1",
   "Condition": {
    "StringEquals": {
     "aws:SourceArn": "arn:aws:s3:::amzn-s3-demo-bucket"
    }
   }
  },
  {
   "Effect": "Allow",
   "Principal": {
    "AWS": "arn:aws:iam::123456789012:user/myUser"
   },
   "Action": "redshift:CreateInboundIntegration",
   "Resource": "arn:aws:redshift-serverless:us-east-1:123456789012:namespace/
namespace-1",
   "Condition": {
    "StringEquals": {
     "aws:SourceArn": "arn:aws:s3:::amzn-s3-demo-bucket"
    }
   }
  }
```

```
 ]
}
```

## Creación de una integración de eventos de S3

Para configurar su trabajo de copia, primero debe definir una integración de eventos de S3.

Amazon Redshift console

Creación de una integración de eventos de Amazon S3 en la consola de Amazon Redshift

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. En el panel de navegación izquierdo, elija Integraciones de eventos de S3.
3. Elija Crear integración de eventos de Amazon S3 para abrir el asistente y crear una integración de eventos de S3 que se utilizará con la copia automática. Su bucket de Amazon S3 de origen y el almacenamiento de datos de Amazon Redshift de destino deben estar en la misma Región de AWS. Especifique la siguiente información al realizar los pasos para crear una integración:

   - Nombre de la integración: es un identificador único en todas las integraciones que son propiedad de su Cuenta de AWS en la Región de AWS actual.
   - Descripción: es un texto que describe la integración de eventos de Amazon S3 para consultarlo más adelante.
   - Bucket de S3 de origen: es el bucket de Amazon S3 en la Cuenta de AWS y la Región de AWS actuales, que es el origen de la ingesta de datos en Amazon Redshift.
   - Almacenamiento de datos de Amazon Redshift: es el clúster aprovisionado de Amazon Redshift de destino o el grupo de trabajo de Redshift sin servidor que recibe los datos de la integración.

     Si su Amazon Redshift objetivo está en la misma cuenta, podrá seleccionar el objetivo. Si el destino está en una cuenta diferente, especifique el ARN del almacenamiento de datos de Amazon Redshift. El destino debe tener una política de recursos con entidades principales y un origen de integración autorizados. Si no tiene las políticas de recursos correctas en el destino y su destino está en la misma cuenta, puede seleccionar la opción Arréglalo por mí para aplicar automáticamente las políticas de recursos durante el proceso de creación de la integración. Si su objetivo se encuentra en otra Cuenta de AWS

diferente, debe aplicar la política de recursos en el almacén de Amazon Redshift de forma manual.

4. Introduzca hasta 50 claves de etiquetas y añada un valor opcional: para proporcionar metadatos adicionales sobre la integración.
5. Se muestra una página de revisión en la que puede seleccionar Crear integración de eventos de S3.

AWS CLI

Para crear una integración de eventos de Amazon S3 mediante la AWS CLI, utilice el comando `create-integration` con las siguientes opciones:

- `integration-name`: especifique un nombre para la integración.
- `source-arn`: especifique el ARN del bucket de origen de Amazon S3.
- `target-arn`: especifique el ARN del espacio de nombres del clúster aprovisionado de Amazon Redshift o del destino del grupo de trabajo de Redshift sin servidor.

El siguiente ejemplo crea una integración proporcionando el nombre de la integración, el ARN de origen y el ARN de destino. La integración no está cifrada.

```
aws redshift create-integration \
--integration-name s3-integration \
--source-arn arn:aws:s3:us-east-1::s3-example-bucket \
--target-arn arn:aws:redshift:us-east-1:123456789012:namespace:a1b2c3d4-5678-90ab-cdef-EXAMPLE22222
          {
    "IntegrationArn": "arn:aws:redshift:us-east-1:123456789012:integration:a1b2c3d4-5678-90ab-cdef-EXAMPLE11111",
    "IntegrationName": "s3-integration",
    "SourceArn": "arn:aws:s3:::s3-example-bucket",
    "SourceType": "s3-event-notifications",
    "TargetArn": "arn:aws:redshift:us-east-1:123456789012:namespace:a1b2c3d4-5678-90ab-cdef-EXAMPLE22222",
    "Status": "creating",
    "Errors": [],
    "CreateTime": "2024-10-09T19:08:52.758000+00:00",
    "Tags": []
}
```

También puede utilizar los siguientes comandos de la AWS CLI para administrar la integración de eventos de S3.

- `delete-integration`: especifique un ARN de integración para eliminar una integración de eventos de S3.
- `modify-integration`: especifique un ARN de integración para cambiar el nombre o la descripción (o ambos) de una integración de eventos de S3.
- `describe-integrations`: especifique un ARN de integración para ver las propiedades de una integración de eventos de S3.

Consulte la Guía de la CLI de Amazon Redshift para obtener más información sobre estos comandos.

A continuación, Amazon Redshift crea una integración de eventos de S3 con el origen y el destino asociados, el estado y la información sobre el estado de un trabajo de copia automática asociado. Para ver la información sobre una integración de eventos de S3 en la consola de Amazon Redshift, seleccione Integraciones de eventos de S3 y elija la integración para ver sus detalles. Hay diferentes integraciones: las que se crean En mi cuenta y las De otras cuentas. La lista En mi cuenta muestra las integraciones en las que el origen y el destino están en la misma cuenta. La lista De otras cuentas muestra las integraciones en las que el origen es propiedad de otra cuenta.

Si elimina una integración de eventos de S3, el estado de COPY JOB correspondiente cambia de 1 (activo) a `0` (inactivo/pendiente). Sin embargo, el COPY JOB correspondiente no se elimina automáticamente. Si más adelante intenta crear un COPY JOB con el mismo nombre, puede que se produzca un conflicto.

## Creación y supervisión de un COPY JOB

Una vez creada la integración, en la página Detalles de la integración de eventos de S3 de la integración que ha creado, elija Crear trabajo de copia automática para ir al editor de consultas V2 de Amazon Redshift, donde puede crear el trabajo de copia automática para la integración. Amazon Redshift hace corresponder el bucket en la cláusula FROM de la instrucción COPY JOB CREATE con el bucket utilizado en la integración de eventos de S3. Para obtener más información acerca del editor de consultas V2 de Amazon Redshift, visite Consulta de una base de datos mediante el editor de consultas V2 en la Guía de administración de Amazon Redshift. Por ejemplo, ejecute el siguiente comando COPY en el editor de consultas V2 para crear un COPY JOB automático que

haga corresponder el bucket de Amazon S3 `s3://amzn-s3-demo-bucket/staging-folder` con una integración de eventos de Amazon S3.

```
COPY public.target_table
FROM 's3://amzn-s3-demo-bucket/staging-folder'
IAM_ROLE 'arn:aws:iam::123456789012:role/MyLoadRoleName'
JOB CREATE my_copy_job_name
AUTO ON;
```

El comando COPY JOB se define una vez. Los mismos parámetros se utilizan para ejecuciones futuras.

Para definir y administrar un COPY JOB, debe tener permiso para hacerlo. Para obtener información sobre cómo conceder y revocar el permiso para un COPY JOB, consulte GRANT y REVOKE. Para obtener información sobre cómo conceder y revocar permisos con ámbito para un COPY JOB, consulte Concesión de permisos acotados y Revocación de los permisos acotados.

Las operaciones de carga se administran utilizando opciones para trabajos CREATE, LIST, SHOW, DROP, ALTER y RUN. Para obtener más información, consulte COPY JOB.

Puede consultar las vistas del sistema para ver el estado y el progreso de COPY JOB. Las vistas se proporcionan de la siguiente manera:

- SYS_COPY_JOB: contiene una fila para cada COPY JOB definido actualmente.
- SYS_COPY_JOB_DETAIL: contiene detalles sobre archivos pendientes, con errores e ingeridos para cada COPY JOB.
- SYS_COPY_JOB_INFO: contiene los mensajes registrados acerca de un COPY JOB.
- SYS_LOAD_HISTORY: contiene detalles de los comandos COPY.
- SYS_LOAD_ERROR_DETAIL: contiene detalles de los errores del comando COPY.
- SVV_COPY_JOB_INTEGRATIONS: contiene detalles de las integraciones de eventos de S3.
- STL_LOAD_ERRORS: contiene errores de los comandos COPY.
- STL_LOAD_COMMITS: contiene información que se utiliza para solucionar problemas de carga de datos del comando COPY.

Para obtener información acerca de la solución de errores de la integración de eventos de S3, consulte Solución de los errores de la integración de eventos de S3 y COPY JOB.

Para obtener la lista de archivos que carga un COPY JOB, ejecute el siguiente SQL, pero primero reemplace *<job_id>*:

```
SELECT job_id, job_name, data_source, copy_query, filename, status, curtime
FROM sys_copy_job copyjob
JOIN stl_load_commits loadcommit
ON copyjob.job_id = loadcommit.copy_job_id
WHERE job_id = <job_id>;
```

## Consideraciones a la hora de crear la integración de eventos de S3 para la copia automática

Tenga en cuenta lo siguiente cuando utilice la copia automática.

- Puede crear un máximo de 200 COPY JOBS para cada clúster o grupo de trabajo de una Cuenta de AWS.
- Puede crear un máximo de 50 integraciones de eventos de S3 para cada destino de Amazon Redshift.
- No puede crear una integración de eventos de S3 con un bucket de Amazon S3 de origen que tenga un punto (.) en su nombre.
- Solo puede crear una integración de eventos de S3 entre origen y destino iguales. Es decir, solo puede haber una integración de eventos de S3 entre un bucket de Amazon S3 y un almacenamiento de datos de Amazon Redshift a la vez.
- No puede tener ninguna notificación de evento existente para el tipo de evento `S3_OBJECT_CREATED` que se define en el bucket de S3 de Amazon de origen. No obstante, una vez creada la integración de un evento de S3, puede actualizar la notificación de eventos del bucket de S3 de Amazon con un prefijo/sufijo de ámbito más restringido. De esta manera, también puede configurar `S3_OBJECT_CREATED` para otro prefijo/sufijo a otros destinos y evitar un conflicto con la integración de eventos de S3. Si experimenta problemas con la copia automática y esta no funciona como esperaba, prepare el registro de AWS CloudTrail de la acción `s3:PutBucketNotificationConfiguration` en el bucket de S3 para el periodo de tiempo en cuestión cuando se ponga en contacto con AWS Support.

# Carga de tablas con comandos DML

Amazon Redshift es compatible con comandos estándar de lenguaje de manipulación de datos (DML) (INSERT, UPDATE y DELETE) que puede usar para modificar las filas de las tablas. También puede utilizar el comando TRUNCATE para realizar eliminaciones masivas rápidas.

> **Note**
>
> Le recomendamos usar el comando COPY para cargar grandes cantidades de datos. El uso de instrucciones INSERT individuales para completar una tabla podría ser prohibitivamente lento. O bien, si los datos ya existen en otras tablas de la base de datos de Amazon Redshift, puede utilizar INSERT INTO… SELECT FROM o CREATE TABLE AS para mejorar el rendimiento. Para obtener información, consulte INSERT o CREATE TABLE AS.

Si inserta, actualiza o elimina una cantidad significativa de filas en una tabla, en relación con la cantidad de filas antes de los cambios, ejecute los comandos ANALYZE y VACUUM en la tabla cuando termine. Si cierta cantidad de cambios pequeños se acumulan en su aplicación con el paso del tiempo, puede que le convenga programar los comandos ANALYZE y VACUUM para que se ejecuten a intervalos regulares. Para obtener más información, consulte Análisis de tablas y Limpieza de tablas.

Temas

- Actualización e inserción de datos nuevos

## Actualización e inserción de datos nuevos

Puede agregar eficazmente nuevos datos a una tabla existente mediante el comando MERGE. Realice una operación de fusión mediante la creación de una tabla provisional y, a continuación, utilice uno de los métodos descritos en esta sección para actualizar la tabla de destino a partir de la tabla provisional. Para obtener más información sobre el comando MERGE, consulte MERGE.

En los Ejemplos de fusión se usa un conjunto de datos de muestra para Amazon Redshift, denominado conjunto de datos TICKIT. Como requisito previo, puede configurar las tablas y los datos de TICKIT según las instrucciones disponibles en Introducción a tareas comunes de bases de datos. Encontrará información más detallada sobre el conjunto de datos de muestra en Base de datos de muestra.

## Método de fusión 1: Reemplazar filas existentes

Si sobrescribe todas las columnas de la tabla de destino, el método más rápido para realizar una combinación es sustituir las filas existentes. Esto escanea la tabla de destino solo una vez, mediante una combinación interna para eliminar las filas que se actualizarán. Una vez eliminadas las filas, se reemplazan por filas nuevas mediante una operación de inserción única desde la tabla provisional.

Aplique este método si se satisfacen todas las siguientes condiciones:

- La tabla destino y la provisional tienen las mismas columnas.
- Desea reemplazar todos los datos de las columnas de la tabla destino por los de las columnas de la tabla provisional.
- Usará todas las filas de la tabla provisional en la fusión.

Si alguno de estos criterios no se aplica, utilice el Método de fusión 2: especificación de una lista de columnas sin utilizar MERGE, descrito en la siguiente sección.

Si no usará todas las filas de la tabla provisional, filtre las instrucciones DELETE e INSERT al usar una cláusula WHERE para excluir las filas que no cambiarán. No obstante, si la mayoría de las filas de la tabla provisional no participarán en la fusión, recomendamos ejecutar los comandos UPDATE e INSERT en distintos pasos, según se describe posteriormente en esta sección.

## Método de fusión 2: especificación de una lista de columnas sin utilizar MERGE

Aplique este método para actualizar columnas específicas de la tabla destino, en lugar de sobrescribir filas enteras. Este método tarda más que el anterior porque requiere un paso de actualización adicional y no utiliza el comando MERGE. Aplique este método si se satisface alguna de las siguientes condiciones:

- No todas las columnas de la tabla destino se actualizarán.
- La mayoría de las filas de la tabla provisional no se usarán en las actualizaciones.

Temas


- Creación de una tabla provisional temporal
- Realización de una operación de fusión al reemplazar filas existentes
- Realización de una operación de fusión al especificar una lista de columnas sin usar el comando MERGE

- Ejemplos de fusión

## Creación de una tabla provisional temporal

La tabla provisional es una tabla temporal que tiene todos los datos que se usarán para hacer los cambios en la tabla de destino, incluidas las actualizaciones e inserciones.

Una operación de fusión requiere una combinación entre la tabla provisional y la tabla destino. Para colocar las filas de combinación, establezca la clave de distribución de la tabla provisional en la misma columna que la clave de distribución de la tabla destino. Por ejemplo, si la tabla destino usa una columna de clave externa como su clave de distribución, use la misma columna para la clave de distribución de la tabla provisional. Si crea la tabla provisional mediante una instrucción CREATE TABLE LIKE, la tabla provisional heredará la clave de distribución de la tabla principal. Si usa una instrucción CREATE TABLE AS, la tabla nueva no hereda la clave de distribución. Para obtener más información, consulte Distribución de datos para la optimización de consultas

Si la clave de distribución no es igual a la clave primaria y si no se actualiza como parte de la operación de fusión, agregue un predicado de combinación redundante en las columnas de la clave de distribución para permitir una combinación colocada. Por ejemplo:

```
where target.primarykey = stage.primarykey
and target.distkey = stage.distkey
```

Para corroborar que la consulta usará una combinación colocada, ejecute la consulta con EXPLAIN y, luego, controle la presencia de DS_DIST_NONE en todas las combinaciones. Para obtener más información, consulte Evaluación del plan de consulta

## Realización de una operación de fusión al reemplazar filas existentes

Cuando ejecute la operación de fusión detallada en el procedimiento, ponga todos los pasos, excepto la creación y eliminación de la tabla provisional, en una transacción única. La transacción se anula si se produce un error en algún paso. El uso de una transacción única también reduce la cantidad de confirmaciones, lo cual ahorra tiempo y recursos.

Pasos para realizar una operación de fusión al reemplazar filas existentes

1. Cree una tabla provisional y, luego, complétela con los datos que se fusionarán, como se muestra en el siguiente pseudocódigo.

```
CREATE temp table stage (like target);

INSERT INTO stage
SELECT * FROM source
WHERE source.filter = 'filter_expression';
```

2. Utilice MERGE para realizar una unión interna con la tabla provisional para actualizar las filas de la tabla de destino que coincidan con la tabla provisional y, a continuación, inserte en la tabla de destino todas las filas restantes que no coincidan con la tabla provisional.

   Le recomendamos que ejecute las operaciones de actualización e inserción en un único comando MERGE.

```
MERGE INTO target
USING stage [optional alias] on (target.primary_key = stage.primary_key)
WHEN MATCHED THEN
UPDATE SET col_name1 = stage.col_name1 , col_name2= stage.col_name2, col_name3 =
 {expr}
WHEN NOT MATCHED THEN
INSERT (col_name1 , col_name2, col_name3) VALUES (stage.col_name1, stage.col_name2,
 {expr});
```

3. Elimine la tabla provisional.

```
DROP TABLE stage;
```

## Realización de una operación de fusión al especificar una lista de columnas sin usar el comando MERGE

Cuando ejecute la operación de fusión detallada en el procedimiento, ponga todos los pasos en una transacción única. La transacción se anula si se produce un error en algún paso. El uso de una transacción única también reduce la cantidad de confirmaciones, lo cual ahorra tiempo y recursos.

Pasos para realizar una operación de fusión al especificar una lista de columnas

1. Coloque toda la operación en un único bloque de transacción.

```
BEGIN transaction;
…
```

```
END transaction;
```

2. Cree una tabla provisional y, luego, complétela con los datos que se fusionarán, como se muestra en el siguiente pseudocódigo.

```
create temp table stage (like target);
insert into stage
select * from source
where source.filter = 'filter_expression';
```

3. Actualice la tabla destino mediante el uso de una combinación interna con la tabla provisional.

   - En la cláusula UPDATE, especifique de manera explícita las columnas que se actualizarán.
   - Realice una combinación interna con la tabla provisional.
   - Si la clave de distribución es diferente a la clave primaria y si no se actualiza, agregue un predicado de combinación redundante en las columnas de la clave de distribución. Para corroborar que la consulta usará una combinación colocada, ejecute la consulta con EXPLAIN y, luego, controle la presencia de DS_DIST_NONE en todas las combinaciones. Para obtener más información, consulte Evaluación del plan de consulta
   - Si la tabla de destino se ordena según la marca temporal, agregue un predicado para aprovechar los análisis de rango restringido en la tabla de destino. Para obtener más información, consulte Prácticas recomendadas de Amazon Redshift para el diseño de consultas.
   - Si no usará todas las filas en la fusión, agregue una cláusula para filtrar las filas que desea cambiar. Por ejemplo, agregue un filtro de desigualdad a una o más columnas para excluir las filas que no han cambiado.
   - Ponga las operaciones de actualización, eliminación e inserción en un bloque de transacción único de manera que, si se produce un problema, todo se revierta.

   Por ejemplo:

```
begin transaction;

update target
set col1 = stage.col1,
col2 = stage.col2,
col3 = 'expression'
from stage
```

```
where target.primarykey = stage.primarykey
and target.distkey = stage.distkey
and target.col3 > 'last_update_time'
and (target.col1 != stage.col1
or target.col2 != stage.col2
or target.col3 = 'filter_expression');
```

4. Elimine las filas innecesarias de la tabla provisional mediante el uso de una combinación interna con la tabla destino. Algunas de las filas de la tabla destino ya coinciden con las filas correspondientes en la tabla provisional y otras se actualizaron en el paso anterior. En cualquiera de los casos, no se necesitan para la inserción.

```
delete from stage
using target
where stage.primarykey = target.primarykey;
```

5. Inserte las filas restantes desde la tabla provisional. Use la misma lista de columnas de la cláusula VALUES que usó en la instrucción UPDATE del paso dos.

```
insert into target
(select col1, col2, 'expression'
from stage);

end transaction;
```

6. Elimine la tabla provisional.

```
drop table stage;
```

## Ejemplos de fusión

En los siguientes ejemplos, se realiza una fusión para actualizar la tabla SALES. En el primer ejemplo, se usa el método más simple de eliminar desde la tabla destino y, luego, se insertan todas las filas desde la tabla provisional. En el segundo ejemplo, se requiere la actualización de columnas seleccionadas de la tabla destino, por lo que se incluye un paso de actualización adicional.

En los Ejemplos de fusión se usa un conjunto de datos de muestra para Amazon Redshift, denominado conjunto de datos TICKIT. Como requisito previo, puede configurar las tablas y los datos de TICKIT siguiendo las instrucciones disponibles en Introducción a las tareas comunes de bases

de datos. Encontrará información más detallada sobre el conjunto de datos de muestra en Base de datos de muestra.

### Origen de datos de la fusión de muestra

Los ejemplos de esta sección requieren de un origen de datos de muestra que incluye tanto actualizaciones como inserciones. Para los ejemplos, crearemos una tabla de muestra denominada SALES_UPDATE que usa datos de la tabla SALES. Llenaremos la tabla nueva con datos aleatorios que representan la actividad de ventas nueva para diciembre. Usaremos la tabla de muestra SALES_UPDATE para crear la tabla provisional de los ejemplos a continuación.

```
-- Create a sample table as a copy of the SALES table.

create table tickit.sales_update as
select * from tickit.sales;

-- Change every fifth row to have updates.

update tickit.sales_update
set qtysold = qtysold*2,
pricepaid = pricepaid*0.8,
commission = commission*1.1
where saletime > '2008-11-30'
and mod(sellerid, 5) = 0;

-- Add some new rows to have inserts.
-- This example creates a duplicate of every fourth row.

insert into tickit.sales_update
select (salesid + 172456) as salesid, listid, sellerid, buyerid, eventid, dateid,
 qtysold, pricepaid, commission, getdate() as saletime
from tickit.sales_update
where saletime > '2008-11-30'
and mod(sellerid, 4) = 0;
```

### Ejemplo de una fusión que reemplaza filas existentes en las claves coincidentes

El siguiente script usa la tabla SALES_UPDATE para realizar una operación de fusión en la tabla SALES con datos nuevos para la actividad de ventas de diciembre. En este ejemplo se reemplazan las filas de la tabla SALES que tienen actualizaciones. Para este ejemplo, actualizaremos las columnas qtysold y pricepaid, pero dejaremos sin cambios commission y saletime.

```
MERGE into tickit.sales
USING tickit.sales_update sales_update
on ( sales.salesid = sales_update.salesid
and sales.listid = sales_update.listid
and sales_update.saletime > '2008-11-30'
and (sales.qtysold != sales_update.qtysold
or sales.pricepaid != sales_update.pricepaid))
WHEN MATCHED THEN
update SET qtysold = sales_update.qtysold,
pricepaid = sales_update.pricepaid
WHEN NOT MATCHED THEN
INSERT (salesid, listid, sellerid, buyerid, eventid, dateid, qtysold , pricepaid,
 commission, saletime)
values (sales_update.salesid, sales_update.listid, sales_update.sellerid,
 sales_update.buyerid, sales_update.eventid,
sales_update.dateid, sales_update.qtysold , sales_update.pricepaid,
 sales_update.commission, sales_update.saletime);

-- Drop the staging table.
drop table tickit.sales_update;

-- Test to see that commission and salestime were not impacted.
SELECT sales.salesid, sales.commission, sales.salestime, sales_update.commission,
 sales_update.salestime
FROM tickit.sales
INNER JOIN tickit.sales_update sales_update
ON
sales.salesid = sales_update.salesid
AND sales.listid = sales_update.listid
AND sales_update.saletime > '2008-11-30'
AND (sales.commission != sales_update.commission
OR sales.salestime != sales_update.salestime);
```

Ejemplo de fusión que especifica una lista de columnas sin utilizar MERGE

En el siguiente ejemplo, se realiza una operación de fusión para actualizar la tabla SALES con datos nuevos para la actividad de ventas de diciembre. Necesitamos datos de muestra que incluyan tanto actualizaciones como inserciones, junto con filas que no hayan cambiado. Para este ejemplo, deseamos actualizar las columnas QTYSOLD y PRICEPAID, y dejar COMMISSION y SALETIME sin cambios. El siguiente script usa la tabla SALES_UPDATE para realizar una operación de fusión en la tabla SALES.

```
-- Create a staging table and populate it with rows from SALES_UPDATE for Dec
create temp table stagesales as select * from sales_update
where saletime > '2008-11-30';

-- Start a new transaction
begin transaction;

-- Update the target table using an inner join with the staging table
-- The join includes a redundant predicate to collocate on the distribution key -- A
 filter on saletime enables a range-restricted scan on SALES

update sales
set qtysold = stagesales.qtysold,
pricepaid = stagesales.pricepaid
from stagesales
where sales.salesid = stagesales.salesid
and sales.listid = stagesales.listid
and stagesales.saletime > '2008-11-30'
and (sales.qtysold != stagesales.qtysold
or sales.pricepaid != stagesales.pricepaid);

-- Delete matching rows from the staging table
-- using an inner join with the target table

delete from stagesales
using sales
where sales.salesid = stagesales.salesid
and sales.listid = stagesales.listid;

-- Insert the remaining rows from the staging table into the target table
insert into sales
select * from stagesales;

-- End transaction and commit
end transaction;

-- Drop the staging table
drop table stagesales;
```

# Realización de una copia profunda

Una copia profunda recrea y vuelve a completar una tabla mediante una operación de inserción masiva, que ordena la tabla de manera automática. Si una tabla tiene una gran región no ordenada, una copia profunda es mucho más rápida que una limpieza. Le recomendamos que solo realice actualizaciones simultáneas durante una operación de copia profunda si puede realizar un seguimiento de ellas. Una vez finalizado el proceso, mueva las actualizaciones delta a la nueva tabla. La operación VACUUM admite de forma automática las actualizaciones simultáneas.

Para crear una copia de una tabla original, puede elegir uno de los siguientes métodos:

- Usar el DDL de la tabla original.

  Si el CREATE TABLE DDL se encuentra disponible, este es el método más rápido y preferido. Si crea una tabla nueva, puede especificar todos los atributos de tabla y columna, entre ellos la clave principal y las claves externas. Para buscar el DDL original, use la función SHOW TABLE.
- Use CREATE TABLE LIKE.

  Si el DDL original no se encuentra disponible, puede usar CREATE TABLE LIKE para recrear la tabla original. La tabla nueva hereda los atributos ENCODING, DISTKEY, SORTKEY y NOT NULL de la tabla principal. La tabla nueva no hereda los atributos de clave principal y clave externa de la tabla principal, pero puede agregarlos con ALTER TABLE.
- Crear una tabla temporal y truncar la tabla original.

  Si debe retener los atributos de clave principal y clave externa de la tabla principal. Si la tabla principal tiene dependencias, puede usar CREATE TABLE... AS (CTAS) para crear una tabla temporal. A continuación, trunque la tabla original y rellénela desde la tabla temporal.

  El uso de una tabla temporal mejora el rendimiento significativamente en comparación con el uso de una tabla permanente, pero existe el riesgo de perder datos. Una tabla temporal se elimina automáticamente al final de la sesión en la que se crea. TRUNCATE se confirma inmediatamente, incluso si se encuentra dentro de un bloque de transacciones. Si TRUNCATE tiene éxito pero la sesión se cierra antes de que se complete la acción INSERT siguiente, los datos se pierden. Si la pérdida de los datos es inaceptable, use una tabla permanente.

Después de crear una copia de una tabla, es posible que deba conceder acceso a la nueva tabla. Puede usar GRANT para definir privilegios de acceso. Para ver y adjudicar todos los privilegios de acceso de una tabla, debe ser uno de los siguientes:

- Un superusuario.
- El propietario de la tabla que desea copiar.
- Un usuario con el privilegio ACCESS SYSTEM TABLE para ver los privilegios de la tabla y con el privilegio de conceder todos los permisos pertinentes.

Además, es posible que tenga que conceder permisos de uso para el esquema en el que se encuentra la copia profunda. Es necesario conceder el permiso de uso si el esquema de la copia profunda es diferente del esquema de la tabla original y también lo es del esquema `public`. Para ver y adjudicar privilegios de uso, debe ser uno de los siguientes:

- Un superusuario.
- Un usuario que puede adjudicar el permiso USAGE para el esquema de la copia profunda.

Pasos para realizar una copia profunda al usar el DDL de la tabla original

1. (Opcional) Recree el DDL de la tabla al ejecutar un script denominado `v_generate_tbl_ddl`.
2. Cree una copia de la tabla al usar el CREATE TABLE DDL original.
3. Use una instrucción INSERT INTO … SELECT para completar la copia con datos desde la tabla original.
4. Compruebe los permisos adjudicados en la tabla anterior. Puede ver estos permisos en la vista de sistema SVV_RELATION_PRIVILEGES.
5. Si es necesario, adjudique los permisos de la tabla antigua a la nueva.
6. Adjudique permisos de uso a todos los grupos y usuarios que tengan privilegios en la tabla original. Este paso no es necesario si la tabla de copia profunda está en el esquema `public` o está en el mismo esquema que la tabla original.
7. Elimine la tabla original.
8. Use una instrucción ALTER TABLE para renombrar la copia con el nombre de la tabla original.

El ejemplo siguiente realiza una copia profunda en la tabla SAMPLE con un duplicado de SAMPLE denominado sample_copy.

```
--Create a copy of the original table in the sample_namespace namespace using the
 original CREATE TABLE DDL.
create table sample_namespace.sample_copy ( … );
```

```
--Populate the copy with data from the original table in the public namespace.
insert into sample_namespace.sample_copy (select * from public.sample);

--Check SVV_RELATION_PRIVILEGES for the original table's privileges.
select * from svv_relation_privileges where namespace_name = 'public' and relation_name
 = 'sample' order by identity_type, identity_id, privilege_type;

--Grant the original table's privileges to the copy table.
grant DELETE on table sample_namespace.sample_copy to group group1;
grant INSERT, UPDATE on table sample_namespace.sample_copy to group group2;
grant SELECT on table sample_namespace.sample_copy to user1;
grant INSERT, SELECT, UPDATE on table sample_namespace.sample_copy to user2;

--Grant usage permission to every group and user that has privileges in the original
 table.
grant USAGE on schema sample_namespace to group group1, group group2, user1, user2;

--Drop the original table.
drop table public.sample;

--Rename the copy table to match the original table's name.
alter table sample_namespace.sample_copy rename to sample;
```

Pasos para realizar una copia profunda al usar CREATE TABLE LIKE

1. Cree una tabla nueva al usar CREATE TABLE LIKE.
2. Use una instrucción INSERT INTO … SELECT para copiar las filas de la tabla actual a la tabla nueva.
3. Compruebe los permisos adjudicados en la tabla anterior. Puede ver estos permisos en la vista de sistema SVV_RELATION_PRIVILEGES.
4. Si es necesario, adjudique los permisos de la tabla antigua a la nueva.
5. Adjudique permisos de uso a todos los grupos y usuarios que tengan privilegios en la tabla original. Este paso no es necesario si la tabla de copia profunda está en el esquema `public` o está en el mismo esquema que la tabla original.
6. Elimine la tabla actual.
7. Use una instrucción ALTER TABLE para renombrar la tabla nueva con el nombre de la tabla original.

En el siguiente ejemplo, se usa CREATE TABLE LIKE para realizar una copia profunda de la tabla SAMPLE.

```
--Create a copy of the original table in the sample_namespace namespace using CREATE
 TABLE LIKE.
create table sameple_namespace.sample_copy (like public.sample);

--Populate the copy with data from the original table.
insert into sample_namespace.sample_copy (select * from public.sample);

--Check SVV_RELATION_PRIVILEGES for the original table's privileges.
select * from svv_relation_privileges where namespace_name = 'public' and relation_name
 = 'sample' order by identity_type, identity_id, privilege_type;

--Grant the original table's privileges to the copy table.
grant DELETE on table sample_namespace.sample_copy to group group1;
grant INSERT, UPDATE on table sample_namespace.sample_copy to group group2;
grant SELECT on table sample_namespace.sample_copy to user1;
grant INSERT, SELECT, UPDATE on table sample_namespace.sample_copy to user2;

--Grant usage permission to every group and user that has privileges in the original
 table.
grant USAGE on schema sample_namespace to group group1, group group2, user1, user2;

--Drop the original table.
drop table public.sample;

--Rename the copy table to match the original table's name.
alter table sample_namespace.sample_copy rename to sample;
```

Pasos para realizar una copia profunda al crear una tabla temporal y truncar la tabla original

1. Use CREATE TABLE AS para crear una tabla temporal con las filas de la tabla original.
2. Trunque la tabla actual.
3. Use una instrucción INSERT INTO … SELECT para copiar las filas de la tabla temporal a la tabla original.
4. Elimine la tabla temporal.

En el siguiente ejemplo, se realiza una copia profunda de la tabla SALES al crear una tabla temporal y truncar la tabla original. Como la tabla original permanece, no es necesario adjudicar permisos a la tabla de copia.

```
--Create a temp table copy using CREATE TABLE AS.
create temp table salestemp as select * from sales;

--Truncate the original table.
truncate sales;

--Copy the rows from the temporary table to the original table.
insert into sales (select * from salestemp);

--Drop the temporary table.
drop table salestemp;
```

# Análisis de tablas

La operación ANALYZE actualiza los metadatos estadísticos que el planificador de consultas usa para seleccionar planes óptimos.

En la mayoría de los casos, no necesita ejecutar de forma explícita el comando ANALYZE. Amazon Redshift monitorea los cambios realizados en la carga de trabajo y, de forma automática, actualiza las estadísticas en segundo plano. Además, el comando COPY realiza un análisis automáticamente cuando carga datos en una tabla vacía.

Para analizar una tabla o toda la base de datos de manera explícita, ejecute el comando ANALYZE.

## Análisis automático

Amazon Redshift monitorea constantemente la base de datos y, de forma automática, realiza operaciones de análisis en segundo plano. Para minimizar el impacto en el rendimiento del sistema, el análisis automático se ejecuta durante periodos en los que hay poca carga de trabajo.

El análisis automático está habilitado de forma predeterminada. Para desactivar el análisis automático, establezca el parámetro `auto_analyze` en **`false`** modificando el grupo de parámetros del clúster.

Para reducir el tiempo de procesamiento y mejorar el rendimiento general del sistema, Amazon Redshift omite el análisis automático de cualquier tabla con pocas modificaciones.

Una operación de análisis omite las tablas que tienen estadísticas actualizadas. Si ejecuta ANALYZE como parte del flujo de trabajo de extracción, transformación y carga (ETL), el análisis automático omite las tablas con estadísticas actuales. De forma similar, un comando ANALYZE explícito omite tablas cuando el análisis automático ha actualizado las estadísticas de la tabla.

## Análisis de los datos nuevos de las tablas

De manera predeterminada, el comando COPY realiza una operación ANALYZE una vez que carga datos en una tabla vacía. Puede forzar una operación ANALYZE sin importar si una tabla se encuentra vacía estableciendo STATUPDATE en ON. Si especifica STATUPDATE en OFF, no se realiza la operación ANALYZE. Solo el propietario de la tabla o un super usuario pueden ejecutar el comando ANALYZE o el comando COPY con STATUPDATE establecido en ON.

Amazon Redshift también analiza las tablas nuevas que se crean con los siguientes comandos:

- CREATE TABLE AS (CTAS)
- CREATE TEMP TABLE AS
- SELECT INTO

Amazon Redshift devuelve un mensaje de advertencia si ejecuta una consulta en una tabla nueva que no se analizó después de la carga de datos inicial. No se producen avisos al ejecutar una consulta en una tabla tras una carga o actualización posterior. El mismo mensaje de aviso se devuelve al ejecutar el comando EXPLAIN en una consulta que hace referencia a tablas que no se han analizado.

Si al agregar datos a una tabla que no está vacía cambia significativamente el tamaño de la tabla, puede actualizar de forma explícita las estadísticas. Para ello, debe ejecutar un comando ANALYZE o usar la opción STATUPDATE ON con el comando COPY. Para ver los detalles de la cantidad de filas que se han insertado o eliminado desde la última vez que se ejecutó ANALYZE, ejecute una consulta en la tabla de catálogo del sistema PG_STATISTIC_INDICATOR.

Puede especificar el ámbito del comando ANALYZE en uno de los siguientes lugares:

- la base de datos actual completa;
- una sola tabla;
- una o más columnas específicas en una tabla única;
- las columnas que, probablemente, se usen como predicados en consultas.

El comando ANALYZE obtiene una muestra de filas desde la tabla, realiza cálculos y guarda las estadísticas de columna resultantes. De manera predeterminada, Amazon Redshift ejecuta un pase de muestra en la columna DISTKEY y otro pase de muestra en todas las demás columnas de la tabla. Si desea generar estadísticas para un subconjunto de columnas, puede especificar una lista de nombres de columnas separada por comas. Puede ejecutar ANALYZE con la cláusula PREDICATE COLUMNS para omitir las columnas que no se usan como predicados.

Las operaciones ANALYZE consumen muchos recursos, por lo que debe ejecutarlas solamente en tablas y columnas que realmente requieran actualizaciones de las estadísticas. No es necesario que analice todas las columnas de las tablas de manera regular ni en el mismo programa. Si los datos cambian considerablemente, analice las columnas que se usan con frecuencia en los siguientes:

- operaciones de agrupamiento y ordenación;
- Uniones
- predicados de consultas.

Para reducir el tiempo de procesamiento y mejorar el rendimiento general del sistema, Amazon Redshift omite ANALYZE en cualquier tabla que tenga un porcentaje bajo de filas modificadas, en función de lo que determine el parámetro analyze_threshold_percent. De manera predeterminada, el umbral de análisis está establecido en 10 por ciento. Puede cambiar el umbral de análisis para la sesión actual al ejecutar un comando SET.

Es menos probable que las columnas que representan hechos, medidas y cualquier atributo relacionado que en realidad nunca se consulta, como las columnas VARCHAR grandes, requieran análisis frecuentes. Por ejemplo, veamos el caso de la tabla LISTING en la base de datos TICKIT.

```
select "column", type, encoding, distkey, sortkey
from pg_table_def where tablename = 'listing';


column         |        type        | encoding | distkey | sortkey
---------------+--------------------+----------+---------+---------
listid         | integer            | none     | t       | 1
sellerid       | integer            | none     | f       | 0
eventid        | integer            | mostly16 | f       | 0
dateid         | smallint           | none     | f       | 0
numtickets     | smallint           | mostly8  | f       | 0
priceperticket | numeric(8,2)       | bytedict | f       | 0
totalprice     | numeric(8,2)       | mostly32 | f       | 0
```

```
listtime       | timestamp with...  | none     | f       | 0
```

Si esta tabla se carga diariamente con una gran cantidad de registros nuevos, la columna LISTID, que frecuentemente se usa en consultas como una clave de combinación, se debe analizar con regularidad. Si TOTALPRICE y LISTTIME se usan con frecuencia como restricciones en consultas, puede analizar esas columnas y la clave de distribución todos los días de la semana.

```
analyze listing(listid, totalprice, listtime);
```

Supongamos que los vendedores y los eventos en la aplicación son mucho más estáticos y los ID de fecha se refieren a un conjunto de días fijo que cubre solo dos o tres años. En este caso, los valores únicos de estas columnas no cambian de manera significativa. No obstante, la cantidad de instancias de cada valor único aumentará de manera constante.

Además, veamos el caso en que se ejecutan consultas sobre las medidas NUMTICKETS y PRICEPERTICKET de manera poco frecuente, en comparación con la columna TOTALPRICE. En este caso, puede ejecutar el comando ANALYZE sobre la tabla completa una vez por fin de semana para actualizar las estadísticas correspondientes a las cinco columnas que no se analizan a diario:

Columnas de predicados

Como alternativa conveniente a especificar una lista de columnas, puede optar por analizar solo las columnas que probablemente utilice como predicados. Al ejecutar una consulta, cualquier columna usada en una combinación, condición de filtro o cláusula de agrupación se marca como columna de predicado en el catálogo de sistema. Al ejecutar ANALYZE con la cláusula PREDICATE COLUMNS, la operación de análisis incluye solo las columnas que cumplen con los siguientes criterios:

- La columna se marca como columna de predicado.
- La columna es una clave de distribución.
- La columna forma parte de una clave de ordenación.

Si ninguna de las columnas de una tabla se marca como predicado, ANALYZE incluye todas las columnas, incluso si se especifica PREDICATE COLUMNS. Si no se marca ninguna columna como columna de predicado, podría ser porque la tabla aún no se ha consultado.

Podría elegir usar PREDICATE COLUMNS una vez que el patrón de consultas de su carga de trabajo esté relativamente estable. Cuando el patrón de consultas es variable, con el uso frecuente de columnas diferentes como predicados, el uso de PREDICATE COLUMNS podría dar como

resultado temporario estadísticas obsoletas. Las estadísticas obsoletas pueden llevar a planes de tiempo de ejecución de consultas poco óptimos y tiempos de ejecución prolongados. No obstante, la próxima vez que ejecute ANALYZE y use PREDICATE COLUMNS, se incluirán las columnas de predicado nuevas.

Para ver detalles de las columnas de predicado, use el siguiente SQL para crear una vista denominada PREDICATE_COLUMNS.

```
CREATE VIEW predicate_columns AS
WITH predicate_column_info as (
SELECT ns.nspname AS schema_name, c.relname AS table_name, a.attnum as col_num,
 a.attname as col_name,
        CASE
            WHEN 10002 = s.stakind1 THEN array_to_string(stavalues1, '||')
            WHEN 10002 = s.stakind2 THEN array_to_string(stavalues2, '||')
            WHEN 10002 = s.stakind3 THEN array_to_string(stavalues3, '||')
            WHEN 10002 = s.stakind4 THEN array_to_string(stavalues4, '||')
            ELSE NULL::varchar
        END AS pred_ts
   FROM pg_statistic s
   JOIN pg_class c ON c.oid = s.starelid
   JOIN pg_namespace ns ON c.relnamespace = ns.oid
   JOIN pg_attribute a ON c.oid = a.attrelid AND a.attnum = s.staattnum)
SELECT schema_name, table_name, col_num, col_name,
       pred_ts NOT LIKE '2000-01-01%' AS is_predicate,
       CASE WHEN pred_ts NOT LIKE '2000-01-01%' THEN (split_part(pred_ts,
 '||',1))::timestamp ELSE NULL::timestamp END as first_predicate_use,
       CASE WHEN pred_ts NOT LIKE '%||2000-01-01%' THEN (split_part(pred_ts,
 '||',2))::timestamp ELSE NULL::timestamp END as last_analyze
FROM predicate_column_info;
```

Supongamos que ejecuta la siguiente consulta en la tabla LISTING. Observe que LISTID, LISTTIME y EVENTID se usan en la combinación, el filtro y las cláusulas de agrupamiento.

```
select s.buyerid,l.eventid, sum(l.totalprice)
from listing l
join sales s on l.listid = s.listid
where l.listtime > '2008-12-01'
group by l.eventid, s.buyerid;
```

Al ejecutar consultas en la vista PREDICATE_COLUMNS, como se observa en el siguiente ejemplo, se ve que LISTID, EVENTID y LISTTIME se marcan como columnas de predicado.

```
select * from predicate_columns
where table_name = 'listing';
```

```
schema_name | table_name | col_num | col_name       | is_predicate |
 first_predicate_use | last_analyze
------------+------------+---------+----------------+--------------
+---------------------+--------------------
public      | listing    |       1 | listid         | true         | 2017-05-05
 19:27:59 | 2017-05-03 18:27:41
public      | listing    |       2 | sellerid       | false        |
  | 2017-05-03 18:27:41
public      | listing    |       3 | eventid        | true         | 2017-05-16
 20:54:32 | 2017-05-03 18:27:41
public      | listing    |       4 | dateid         | false        |
  | 2017-05-03 18:27:41
public      | listing    |       5 | numtickets     | false        |
  | 2017-05-03 18:27:41
public      | listing    |       6 | priceperticket | false        |
  | 2017-05-03 18:27:41
public      | listing    |       7 | totalprice     | false        |
  | 2017-05-03 18:27:41
public      | listing    |       8 | listtime       | true         | 2017-05-16
 20:54:32 | 2017-05-03 18:27:41
```

Mantener las estadísticas actualizadas mejora el rendimiento de las consultas, ya que permite que el planificador de consultas elija los planes óptimos. Amazon Redshift actualiza las estadísticas de forma automática en segundo plano, y usted también puede ejecutar de forma explícita el comando ANALYZE. Si decide ejecutar explícitamente el comando ANALYZE, haga lo siguiente:

- Ejecute el comando ANALYZE antes de ejecutar cualquier consulta.
- Ejecute el comando ANALYZE en la base de datos de manera habitual al final de cada ciclo de carga o actualización regular.
- Ejecute el comando ANALYZE en toda tabla nueva que cree y en cualquier tabla o columna existente que sufra cambios significativos.
- Considere ejecutar operaciones ANALYZE en programas diferentes para diferentes tipos de tablas y columnas, según su uso en consultas y su tendencia al cambio.
- Para ahorrar tiempo y recursos del clúster, al ejecutar ANALYZE use la cláusula PREDICATE COLUMNS.

No tiene que ejecutar explícitamente el comando ANALYZE después de restaurar una instantánea en un clúster aprovisionado o en un espacio de nombres sin servidor, ni después de reanudar un clúster aprovisionado en pausa. Amazon Redshift conserva la información de la tabla del sistema en estos casos, lo que hace innecesarios los comandos ANALYZE manuales. Amazon Redshift seguirá ejecutando operaciones de análisis automáticas según sea necesario.

Una operación de análisis omite las tablas que tienen estadísticas actualizadas. Si ejecuta ANALYZE como parte del flujo de trabajo de extracción, transformación y carga (ETL), el análisis automático omite las tablas con estadísticas actuales. De forma similar, un comando ANALYZE explícito omite tablas cuando el análisis automático ha actualizado las estadísticas de la tabla.

## Historial del comando ANALYZE

Saber cuándo fue la última vez que el comando ANALYZE se ejecutó en una tabla o base de datos es útil. Cuando se ejecuta el comando ANALYZE, Amazon Redshift ejecuta distintas consultas que se ven de la siguiente manera:

```
padb_fetch_sample: select * from table_name
```

Ejecute la consulta STL_ANALYZE para ver el historial de las operaciones ANALYZE. Si Amazon Redshift analiza una tabla mediante el análisis automático, la columna `is_background` se establece en `t` (true). De lo contrario, se establece en `f` (falso). En el siguiente ejemplo, se combina STV_TBL_PERM para mostrar el nombre de la tabla y los detalles de tiempo de ejecución.

```
select distinct a.xid, trim(t.name) as name, a.status, a.rows, a.modified_rows,
 a.starttime, a.endtime
from stl_analyze a
join stv_tbl_perm t  on t.id=a.table_id
where name = 'users'
order by starttime;


xid    | name  | status          | rows  | modified_rows | starttime           |
 endtime
-------+-------+-----------------+-------+---------------+---------------------
+---------------------
  1582 | users | Full            | 49990 |         49990 | 2016-09-22 22:02:23 |
 2016-09-22 22:02:28
244287 | users | Full            | 24992 |         74988 | 2016-10-04 22:50:58 |
 2016-10-04 22:51:01
```

```
244712 | users | Full            | 49984 |         24992 | 2016-10-04 22:56:07 |
 2016-10-04 22:56:07
245071 | users | Skipped         | 49984 |             0 | 2016-10-04 22:58:17 |
 2016-10-04 22:58:17
245439 | users | Skipped         | 49984 |          1982 | 2016-10-04 23:00:13 |
 2016-10-04 23:00:13
(5 rows)
```

De manera alternativa, puede ejecutar una consulta más compleja que devuelva todas las instrucciones que se ejecutaron en cada transacción completada que incluyó un comando ANALYZE:

```
select xid, to_char(starttime, 'HH24:MM:SS.MS') as starttime,
datediff(sec,starttime,endtime ) as secs, substring(text, 1, 40)
from svl_statementtext
where sequence = 0
and xid in (select xid from svl_statementtext s where s.text like 'padb_fetch_sample
%' )
order by xid desc, starttime;

xid  |  starttime   | secs |                 substring
-----+--------------+------+-----------------------------------------
1338 | 12:04:28.511 |    4 | Analyze date
1338 | 12:04:28.511 |    1 | padb_fetch_sample: select count(*) from
1338 | 12:04:29.443 |    2 | padb_fetch_sample: select * from date
1338 | 12:04:31.456 |    1 | padb_fetch_sample: select * from date
1337 | 12:04:24.388 |    1 | padb_fetch_sample: select count(*) from
1337 | 12:04:24.388 |    4 | Analyze sales
1337 | 12:04:25.322 |    2 | padb_fetch_sample: select * from sales
1337 | 12:04:27.363 |    1 | padb_fetch_sample: select * from sales
...
```

# Limpieza de tablas

Amazon Redshift puede ordenar y realizar una operación VACUUM DELETE de forma automática en las tablas en segundo plano. Para limpiar las tablas tras una carga o una serie de actualizaciones incrementales, también puede ejecutar el comando VACUUM, ya sea en la base de datos completa o en tablas individuales.

> **Note**
>
> Solo los usuarios con los permisos de tabla necesarios pueden vaciar una tabla de forma eficaz. Si se ejecuta VACUUM sin los privilegios de tabla necesarios, la operación se completa correctamente pero no tiene ningún efecto. Para obtener una lista de los permisos de tabla válidos para ejecutar VACUUM de forma eficaz, consulte VACUUM.
> Por esta razón, recomendamos realizar una limpieza de tablas individual según sea necesario. También recomendamos esta opción porque es probable que una limpieza de la base de datos completa sea una operación costosa.

## Clasificación automática de tablas

Amazon Redshift ordena los datos en segundo plano de forma automática para mantener los datos de la tabla en el orden de su clave de ordenación. Amazon Redshift realiza un seguimiento de sus consultas de análisis para determinar las secciones de la tabla que se beneficiarán de la ordenación.

Según la carga en el sistema, Amazon Redshift inicia de forma automática la ordenación. La ordenación automática disminuye la necesidad de ejecutar el comando VACUUM para mantener los datos en orden de clave de ordenación. Si necesita que los datos estén ordenados en clave de ordenación, por ejemplo después de una gran carga de datos, puede ejecutar manualmente el comando VACUUM. Para determinar si su tabla se beneficia al ejecutar VACUUM SORT, monitorice la columna `vacuum_sort_benefit` en SVV_TABLE_INFO.

Amazon Redshift realiza un seguimiento de las consultas de análisis que utilizan la clave de ordenación en cada tabla. Amazon Redshift calcula el porcentaje máximo de mejora al analizar y filtrar los datos de cada tabla (si la tabla estaba completamente ordenada). Esta estimación está visible en la columna `vacuum_sort_benefit` en SVV_TABLE_INFO. Puede utilizar esta columna junto con la columna `unsorted` para determinar si las consultas pueden beneficiarse de la ejecución manual de VACUUM SORT en una tabla. La columna `unsorted` refleja el orden de clasificación físico de una tabla. La columna `vacuum_sort_benefit` especifica el impacto que tiene ordenar una tabla mediante la ejecución de VACUUM SORT de forma manual.

Por ejemplo, fíjese en la consulta siguiente:

```
select "table", unsorted,vacuum_sort_benefit from svv_table_info order by 1;
```

```
 table | unsorted | vacuum_sort_benefit
-------+----------+---------------------
 sales |    85.71 |                5.00
 event |    45.24 |               67.00
```

Para la tabla “sales”, aunque la tabla está físicamente desordenada un 86 %, la repercusión del rendimiento de la consulta de la tabla estando desordenada un 86 % es de solo un 5 %. Esto puede ser porque solo se puede acceder a una pequeña parte de la tabla por consultas o porque muy pocas consultas accedieron a la tabla. Para la tabla “event”, la tabla está desordenada físicamente un 45 %. Pero la repercusión del rendimiento de la consulta de un 67 % indica que se puede acceder a una gran parte de la tabla por consultas o que la cantidad de consultas que accedieron a la tabla fue grande. La tabla “event” se puede beneficiar potencialmente de ejecutar VACUUM SORT.

## Eliminación de limpieza automática

Cuando se realiza una eliminación, las filas se marcan para eliminarse, pero esto no sucede. Amazon Redshift ejecuta de forma automática una operación VACUUM DELETE en segundo plano en función de la cantidad de filas eliminadas en las tablas de base de datos. Amazon Redshift programa la operación VACUUM DELETE para que se ejecute durante periodos de carga reducida y pone en pausa la operación durante periodos de carga elevada.

Temas

- Frecuencia de ejecución de VACUUM
- Fase de ordenación y fase de fusión
- Umbral de limpieza
- Tipos de limpieza
- Cómo minimizar los tiempos de limpieza

## Frecuencia de ejecución de VACUUM

Debe limpiar con tanta frecuencia como sea necesario para mantener un rendimiento de consultas coherente. Al determinar con cuánta frecuencia debe ejecutar el comando VACUUM, considere los siguientes factores:

- Ejecute VACUUM durante periodos en los que espera actividad mínima en el clúster, como por la noche o durante ventanas de administración de la base de datos designadas.

- Ejecute los comandos VACUUM fuera de los periodos de mantenimiento. Para obtener más información, consulte Programación de periodos de mantenimiento.
- Una región grande sin ordenar da como resultado tiempos de limpieza más prolongados. Si retrasa la limpieza, esta demorará más porque se necesitará reorganizar más datos.
- VACUUM es una operación con muchas E/S, por lo que cuanto más demore en completarse la limpieza, más efecto tendrá sobre las consultas simultáneas y otras operaciones de base de datos que se ejecuten en su clúster.
- VACUUM demora más en tablas con ordenación intercalada. Para evaluar si las tablas intercaladas necesitan reordenarse, consulte la vista SVV_INTERLEAVED_COLUMNS.

## Fase de ordenación y fase de fusión

Amazon Redshift realiza una operación de limpieza en dos fases: primero, ordena las filas de la región no ordenada y, luego, de ser necesario, fusiona las filas recién ordenadas con las filas ya existentes en el final de la tabla. Al limpiar una tabla grande, la operación de limpieza avanza en una serie de pasos que consisten en ordenaciones incrementales seguidas de fusiones. Si se produce un error en la operación o si Amazon Redshift se desconecta durante una limpieza, la tabla o la base de datos limpiada de forma parcial se encontrarán en estado consistente, pero debe reiniciar la operación de limpieza de forma manual. Las ordenaciones incrementales se pierden, pero las filas fusionadas confirmadas antes del error no necesitar limpiarse de nuevo. Si la región sin ordenar es grande, el tiempo perdido podría ser significativo. Para obtener más información acerca de las fases de ordenación y fusión, consulte Reducción del volumen de filas fusionadas.

Los usuarios pueden obtener acceso a las tablas mientras se limpian. Puede realizar consultas y escribir operaciones mientras se limpia una tabla, pero cuando se ejecutan DML y una limpieza de manera simultánea, ambas operaciones podrían demorar más de lo usual. Si se ejecutan las instrucciones UPDATE y DELETE durante una limpieza, el rendimiento del sistema podría reducirse. Las fusiones incrementales bloquean temporalmente las operaciones UPDATE y DELETE simultáneas, y estas últimas, a su vez, bloquean temporalmente los pasos de fusión incrementales en las tablas afectadas. Las operaciones DDL, como ALTER TABLE, se bloquean hasta que la operación de limpieza finaliza con la tabla.

> **Note**
>
> Hay varios modificadores de VACUUM que controlan la forma en que funciona. Puede utilizarlos para adaptar la operación de limpieza a las necesidades actuales. Por ejemplo,

el uso de VACUUM RECLUSTER acorta la operación de limpieza, ya que no se realiza una operación de fusión completa. Para obtener más información, consulte VACUUM.

## Umbral de limpieza

De manera predeterminada, VACUUM omite la fase de ordenación para cualquier tabla que tenga más del 95 por ciento de las filas de la tabla ordenadas. La omisión de la fase de ordenación puede mejorar significativamente el rendimiento de VACUUM. Para cambiar el umbral de ordenación predeterminado de una tabla única, incluya el nombre de la tabla y el parámetro TO threshold (umbral) PERCENT cuando ejecute el comando VACUUM.

## Tipos de limpieza

Para obtener más información acerca de los distintos tipos de limpieza, consulte VACUUM.

## Cómo minimizar los tiempos de limpieza

Amazon Redshift ordena los datos de forma automática y ejecuta VACUUM DELETE en segundo plano. Esto disminuye la necesidad de ejecutar el comando VACUUM. La limpieza es un proceso que puede durar mucho tiempo. Según la naturaleza de los datos, le recomendamos las siguientes prácticas para minimizar los tiempos de limpieza.

Temas


- Decisión de si se debe reindexar
- Reducción del tamaño de la región no ordenada
- Reducción del volumen de filas fusionadas
- Carga de los datos en orden de clave de clasificación
- Uso de tablas de series temporales para reducir los datos almacenados


### Decisión de si se debe reindexar

A menudo puede mejorar de manera significativa el rendimiento de consultas al usar un estilo de ordenación intercalada, pero, con el tiempo, el rendimiento podría degradarse si cambia la distribución de los valores en las columnas con claves de ordenación.

Cuando inicialmente se carga una tabla intercalada vacía mediante COPY o CREATE TABLE AS, Amazon Redshift crea el índice intercalado de forma automática. Si carga inicialmente una tabla

intercalada mediante INSERT, necesita ejecutar VACUUM REINDEX a continuación para inicializar el índice intercalado.

Con el tiempo, a medida que agrega filas con valores de clave de ordenación nuevos, el rendimiento podría degradarse si cambia la distribución de los valores en las columnas con claves de ordenación. Si las filas nuevas caen principalmente dentro del rango de los valores de clave de ordenación existentes, la reindexación no es necesaria. Ejecute VACUUM SORT ONLY o VACUUM FULL para restaurar el orden de ordenación.

El motor de consultas puede usar el orden de ordenación para seleccionar de manera eficiente qué bloques de datos deben analizarse para procesar una consulta. En caso de una ordenación intercalada, Amazon Redshift analiza los valores en las columnas con clave de ordenación para determinar el orden óptimo. Si la distribución de los valores de clave cambia o se sesga a medida que se agregan filas, la estrategia de ordenación ya no será óptima y el beneficio en rendimiento brindado por la ordenación se degradará. Para volver a analizar la distribución de claves de ordenación se puede ejecutar VACUUM REINDEX. La operación de reindexación lleva mucho tiempo; en consecuencia, para decidir si una tabla se beneficiará de una reindexación, ejecute una consulta en la vista SVV_INTERLEAVED_COLUMNS.

Por ejemplo, la siguiente consulta muestra los detalles de las tablas que usan claves de ordenación intercalada.

```
select tbl as tbl_id, stv_tbl_perm.name as table_name,
col, interleaved_skew, last_reindex
from svv_interleaved_columns, stv_tbl_perm
where svv_interleaved_columns.tbl = stv_tbl_perm.id
and interleaved_skew is not null;


 tbl_id | table_name | col | interleaved_skew | last_reindex
--------+------------+-----+------------------+---------------------
 100048 | customer   |   0 |             3.65 | 2015-04-22 22:05:45
 100068 | lineorder  |   1 |             2.65 | 2015-04-22 22:05:45
 100072 | part       |   0 |             1.65 | 2015-04-22 22:05:45
 100077 | supplier   |   1 |             1.00 | 2015-04-22 22:05:45
(4 rows)
```

El valor de `interleaved_skew` es una relación que indica el grado de sesgo. Un valor de 1 indica que no existe sesgo. Si el sesgo es mayor que 1,4, VACUUM REINDEX usualmente mejorará el rendimiento a menos que el sesgo en el conjunto subyacente sea inherente.

Puede usar el valor de fecha en `last_reindex` para determinar cuánto tiempo ha pasado desde la última reindexación.

## Reducción del tamaño de la región no ordenada

La región no ordenada aumenta cuando se cargan grandes cantidades de datos nuevos en tablas que ya tienen datos o cuando no se limpian las tablas como parte de las operaciones de mantenimiento de rutina. Para evitar las operaciones de limpieza prolongadas, aplique las siguientes prácticas:

- Ejecutar las operaciones de limpieza según un programa regular.

  Si carga tablas en incrementos pequeños (como actualizaciones diarias que representan un bajo porcentaje de la cantidad total de filas de la tabla), el ejecutar VACUUM con regularidad lo ayudará a asegurarse de que las operaciones de limpieza individuales se realicen con rapidez.
- Ejecutar la carga más grande primero.

  Si necesita cargar una tabla nueva con distintas operaciones COPY, ejecute la carga más grande primero. Cuando se ejecuta una carga inicial en una tabla nueva o truncada, todos los datos se cargan directamente en la región ordenada, por lo que no se requiere limpieza alguna.
- Truncar una tabla en lugar de eliminar todas las filas.

  Eliminar las filas de una tabla no recupera el espacio que estas ocupaban sino hasta que se realiza una operación de limpieza; no obstante, el truncar una tabla la vacía y recupera el espacio en el disco, por lo que no se requiere limpieza alguna. De manera alternativa, elimine la tabla y recréela.
- Truncar o eliminar las tablas de prueba.

  Si carga una cantidad pequeña de filas en una tabla con fines de prueba, no elimine las filas una vez que termine. En cambio, trunque la tabla y vuelva a cargar las filas como parte de la operación de carga de producción subsecuente.
- Realizar una copia profunda.

  Si una tabla que usa una tabla con clave de ordenación compuesta tiene una región no ordenada grande, una copia profunda es mucho más rápida que una limpieza. Una copia profunda recrea y vuelve a completar una tabla mediante una operación de inserción masiva, que vuelve a ordenar la tabla de manera automática. Si una tabla tiene una gran región no ordenada, una copia profunda es mucho más rápida que una limpieza. La diferencia es que durante una operación de copia profunda no se pueden hacer actualizaciones simultáneas, lo cual sí puede hacerse durante una

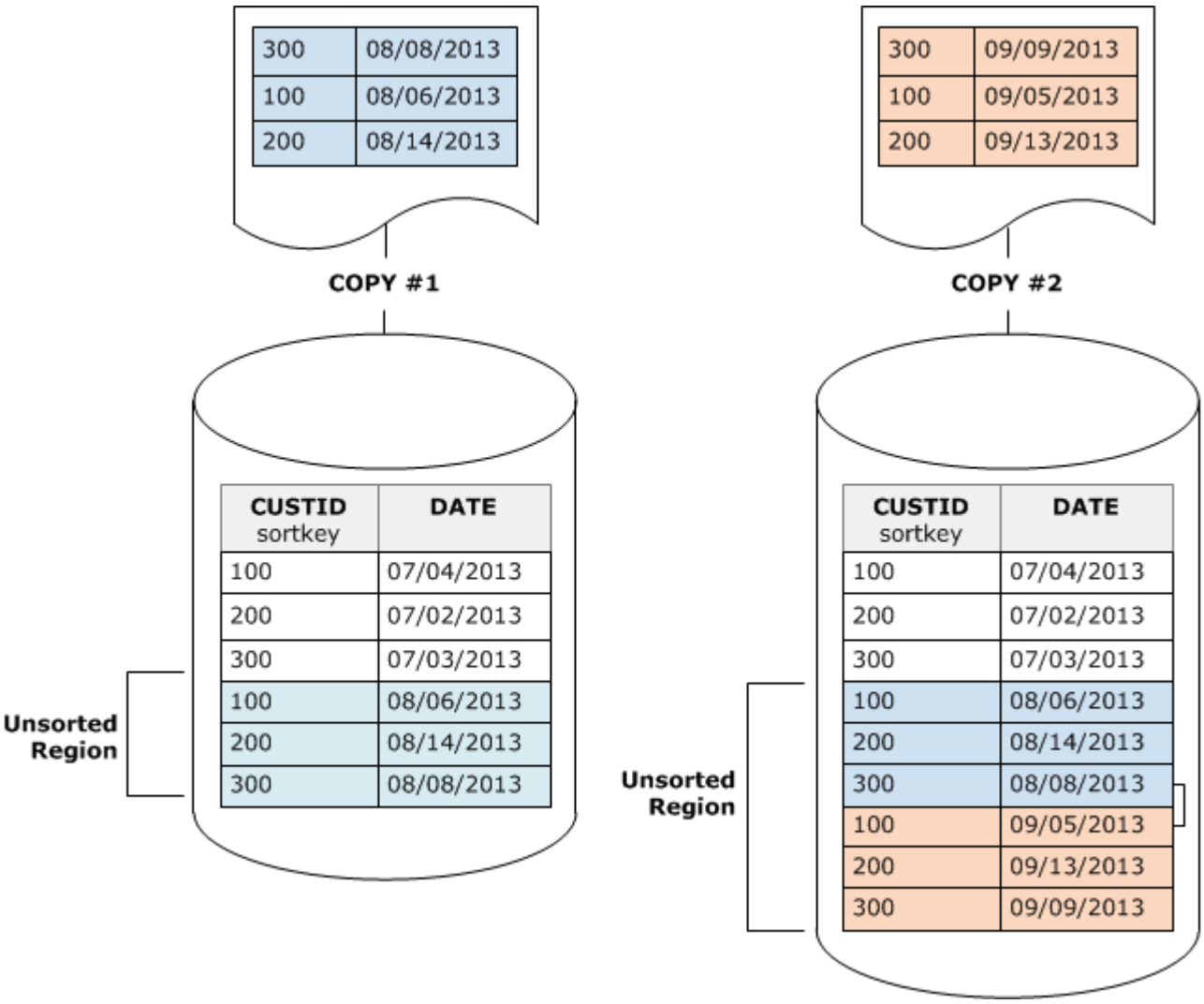
limpieza. Para obtener más información, consulte Prácticas recomendadas de Amazon Redshift para el diseño de consultas.

## Reducción del volumen de filas fusionadas

Si una operación de limpieza necesita fusionar filas nuevas en la región ordenada de una tabla, el tiempo requerido para una limpieza aumentará a medida que la tabla aumente. Puede mejorar el rendimiento de la limpieza al reducir la cantidad de filas que deben fusionarse.

Antes de una limpieza, una tabla consta de una región ordenada en la parte superior, seguida de una región no ordenada que aumenta cada vez que se agregan o actualizan filas. Cuando se agrega un conjunto de filas mediante una operación COPY, el conjunto nuevo de filas se ordena en la clave de ordenación a medida que se agrega a la región no ordenada que se encuentra al final de la tabla. Las filas nuevas se ordenan dentro de su propio conjunto, pero no dentro de la región no ordenada.

En el siguiente diagrama, se ilustra la región no ordenada luego de dos operaciones COPY sucesivas, donde la clave de ordenación es CUSTID. Para mayor simpleza, en este ejemplo se muestra una clave de ordenación compuesta, pero los mismos principios se aplican a las claves de ordenación intercaladas, excepto que, para las tablas intercaladas, el impacto de la región no ordenada es mayor.




Una operación de limpieza restaura el orden de la tabla en dos fases:

1. Ordena la región no ordenada en una región recién ordenada.

   La primera fase es relativamente poco costosa, porque solo se rescribe la región no ordenada. Si el rango de valores de clave de clasificación de la región recién ordenada es mayor que el rango existente, solo se deben rescribir las filas nuevas y la limpieza se completa. Por ejemplo, si la región ordenada tiene valores de ID de 1 a 500 y las operaciones de copia subsecuentes agregan valores de clave mayores que 500, solo se debe rescribir la región no ordenada.

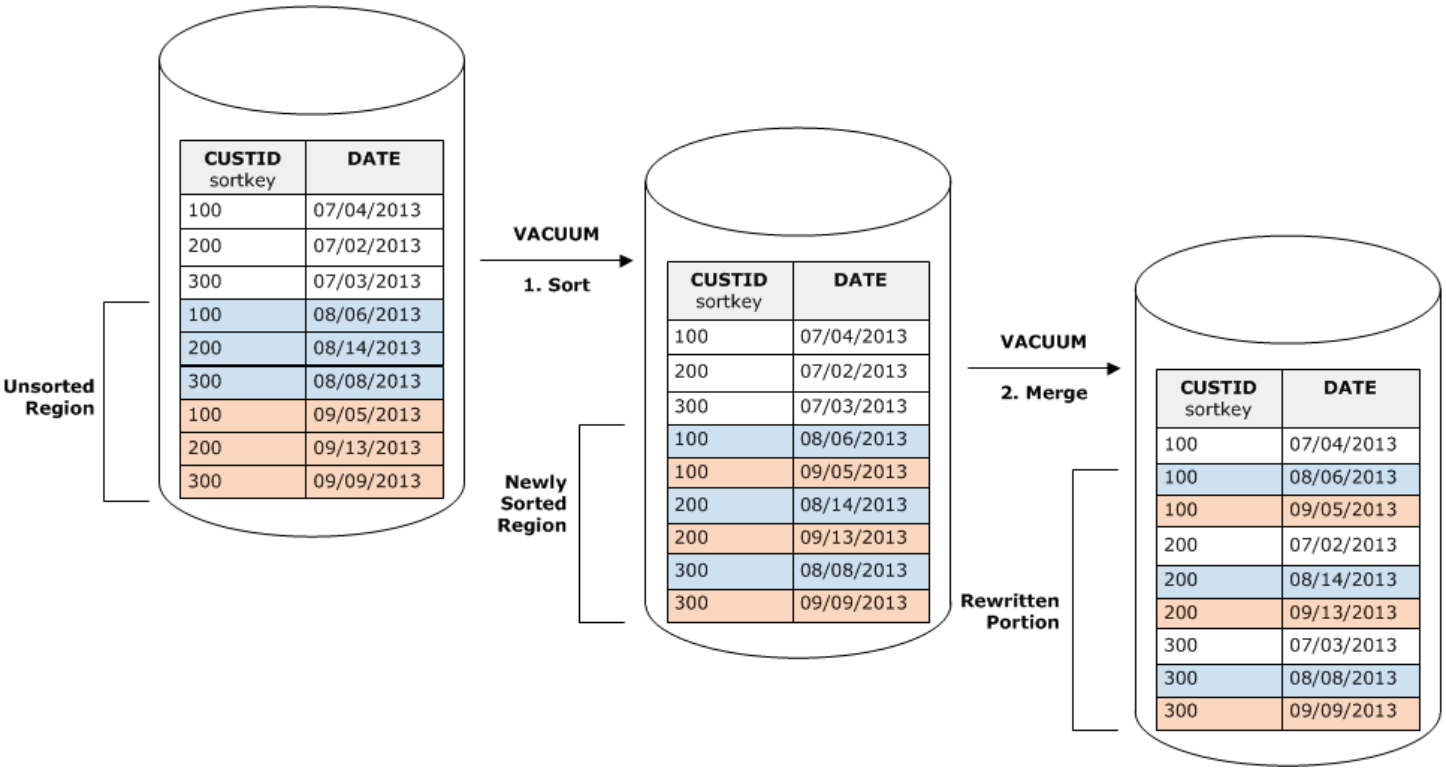
2. Fusiona la región recién ordenada con la región anteriormente ordenada.

   Si las claves de la región recién ordenada se superponen con las claves de la región ordenada, VACUUM necesita fusionar las filas. Desde el comienzo de la región recién ordenada (en la clave

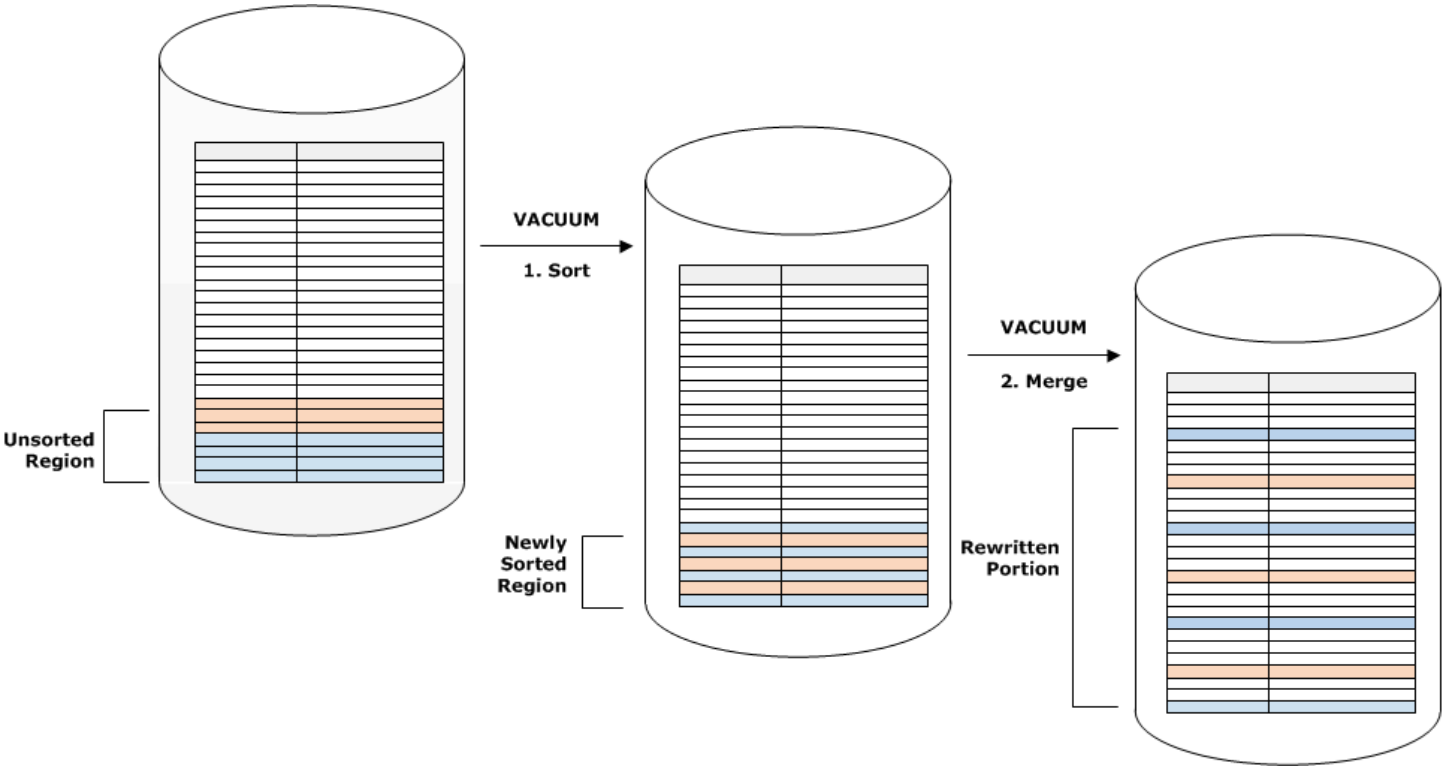
de ordenación más baja), la limpieza escribe las filas fusionadas de las regiones anteriormente y recién ordenadas en un nuevo conjunto de bloques.

El grado de superposición del nuevo rango de claves de ordenación en relación con las claves de ordenación existentes determina hasta dónde es necesario rescribir la región antes ordenada. Si las claves no ordenadas se encuentran esparcidas por todo el rango de ordenación existente, podría necesitarse una limpieza que rescriba partes de la tabla.

En el siguiente diagrama, se muestra cómo una limpieza ordenaría y fusionaría las filas agregadas a una tabla en la que CUSTID es la clave de ordenación. Como cada operación de COPY agrega un nuevo conjunto de filas con valores de clave que se superponen con las claves existentes, se debe rescribir casi toda la tabla. En el diagrama, se muestra una única ordenación y fusión, pero, en la práctica, una operación de limpieza grande consiste en una serie de pasos incrementales de ordenación y fusión.




Si el rango de las claves de ordenación en un conjunto de filas nuevas se superpone con el rango de claves existentes, el costo de la fase de fusión continúa aumentando de manera proporcional al tamaño de la tabla, a medida que la tabla aumenta mientras el costo de la fase de ordenación se mantiene proporcional al tamaño de la región no ordenada. En tal caso, el costo de la fase de fusión supera el costo de la fase de ordenación, como lo muestra el diagrama a continuación.




Para determinar qué proporción de una tabla se volvió a fusionar, ejecute una consulta SVV_VACUUM_SUMMARY una vez terminada la operación de limpieza. La siguiente consulta muestra el efecto de seis limpiezas consecutivas, a medida que CUSTSALES aumenta cada vez más con el paso del tiempo.

```
select * from svv_vacuum_summary
where table_name = 'custsales';


 table_name | xid  | sort_      | merge_     | elapsed_   | row_  | sortedrow_ | block_
  | max_merge_
            |      | partitions | increments | time       | delta | delta      | delta
  | partitions
 -----------+------+------------+------------+------------+-------+------------
+---------+--------------
  custsales | 7072 |          3 |          2 |  143918314 |     0 |   88297472 |   1524
  |      47
  custsales | 7122 |          3 |          3 |  164157882 |     0 |   88297472 |    772
  |      47
  custsales | 7212 |          3 |          4 |  187433171 |     0 |   88297472 |    767
  |      47
  custsales | 7289 |          3 |          4 |  255482945 |     0 |   88297472 |    770
  |      47
```

```
 custsales | 7420 |          3 |          5 |  316583833 |     0 |   88297472 |    769
 |      47
 custsales | 9007 |          3 |          6 |  306685472 |     0 |   88297472 |    772
 |      47
(6 rows)
```

La columna merge_increments indica la cantidad de datos fusionados en cada operación de limpieza. Si la cantidad de incrementos de fusión durante las limpiezas consecutivas aumenta de manera proporcional al crecimiento en el tamaño de la tabla, indica que cada operación de limpieza vuelve a fusionar una cantidad cada vez mayor de filas en la tabla, debido a que las regiones ordenadas ya existentes y las nuevas se superponen.

## Carga de los datos en orden de clave de clasificación

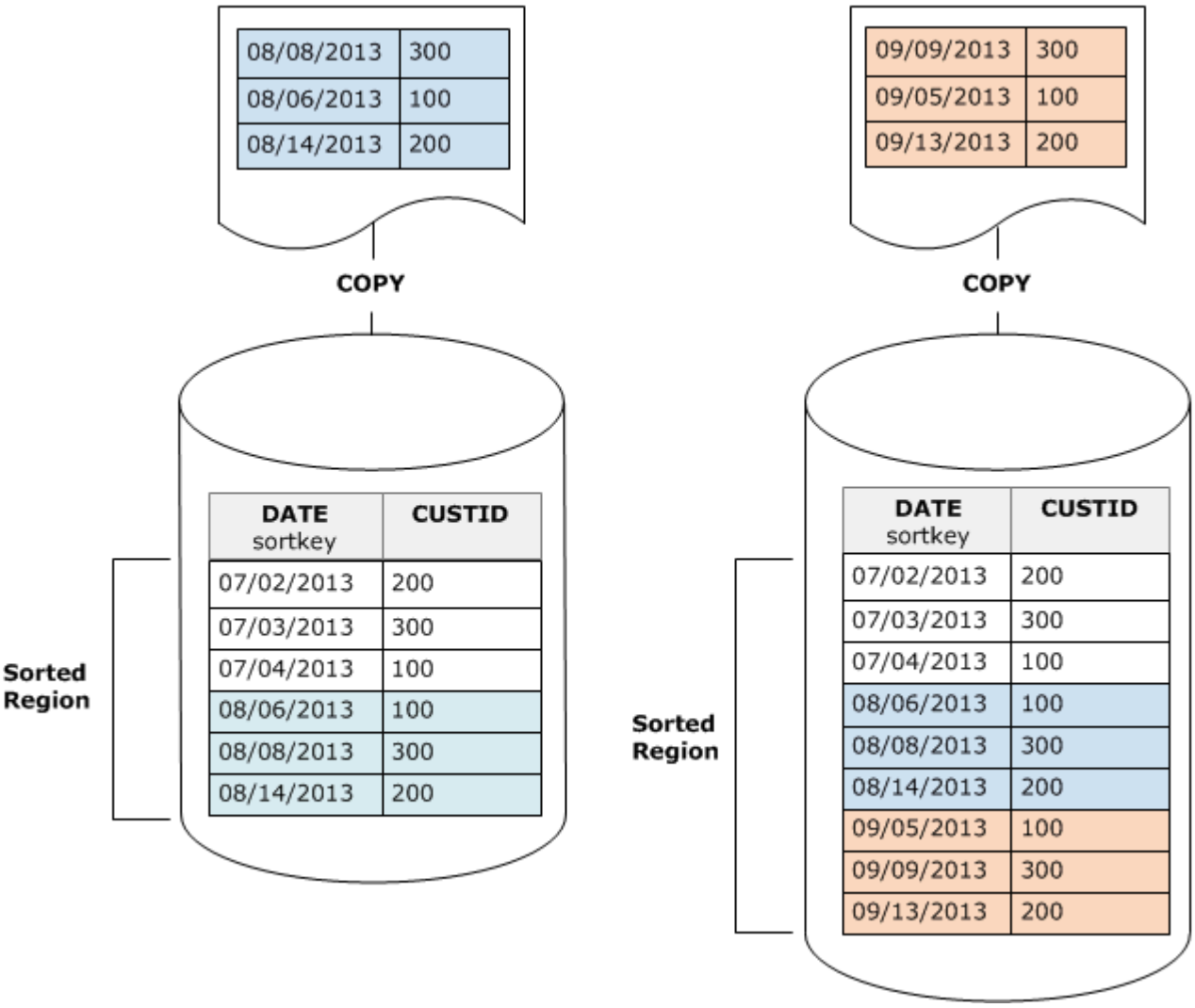
Si carga los datos en orden de clave de clasificación mediante un comando COPY, es posible que reduzca o incluso elimine la necesidad de ejecutar una limpieza.

COPY agrega automáticamente filas nuevas a la región ordenada de la tabla cuando se satisfacen todas las siguientes condiciones:

- La tabla usa una clave de ordenación compuesta con solo una columna de ordenación.
- La columna de ordenación es NOT NULL.
- La tabla está 100 por ciento ordenada o vacía.
- Todas las filas nuevas son mayores en cuanto al orden de ordenación que las filas existentes, entre ellas las filas marcadas para eliminación. En esta instancia, Amazon Redshift usa los primeros ocho bytes de la clave de ordenación para determinar el orden.

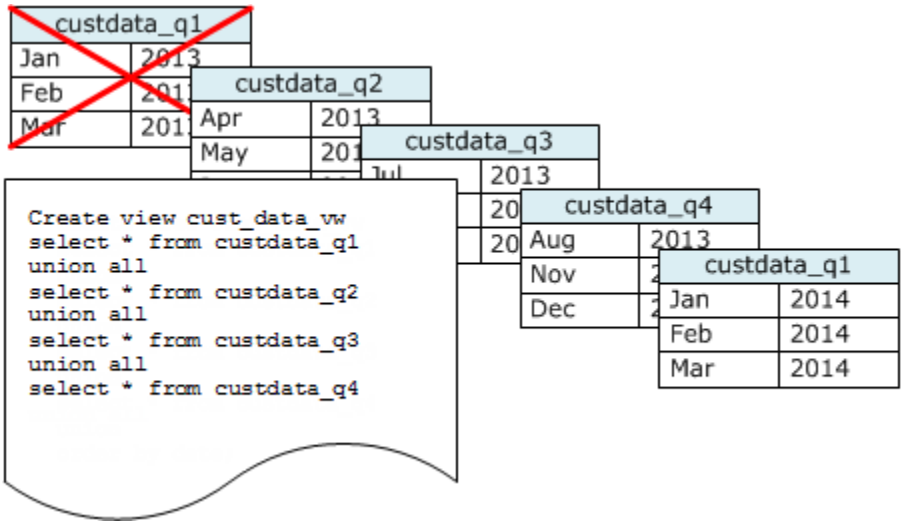
Suponga, por ejemplo, que tiene una tabla que registra los eventos de los clientes con un ID de cliente y la hora. Si ordena a partir del ID de cliente, es probable que el rango de clave de ordenación de las filas nuevas agregadas con cargas incrementales se superponga con el rango existente, como se observa en el ejemplo anterior, lo cual conduce a una operación de limpieza costosa.

Si establece la clave de clasificación en una columna de marca temporal, las filas nuevas se anexarán en orden de ordenación al final de la tabla, como lo muestra el diagrama a continuación, lo que reducirá o incluso eliminará la necesidad de ejecutar una limpieza.




## Uso de tablas de series temporales para reducir los datos almacenados

Si mantiene datos durante un periodo acumulativo, use una serie de tablas, como se ilustra en el siguiente diagrama.




Cree una tabla nueva cada vez que agrega un conjunto de datos y, luego, elimine la tabla más antigua de la serie. Se obtiene un beneficio doble:

- Se evita el costo adicional de eliminar filas, porque una operación DROP TABLE es mucho más eficiente que una operación DELETE masiva.
- Si las tablas se ordenan por marca temporal, no se necesita una limpieza. Si cada tabla tiene los datos correspondientes a un mes, una limpieza deberá, cuanto mucho, rescribir los datos de todo un mes, incluso si las tablas no se ordenan por marca temporal.

Puede crear una vista UNION ALL para usar al informar consultas, lo cual oculta el hecho de que los datos se almacenan en distintas tablas. Si una consulta filtra según la clave de ordenación, el planificador de consultas puede omitir de manera eficiente todas las tablas que no se usan. UNION ALL puede ser menos eficiente para otros tipos de consultas, por lo que debería evaluar el rendimiento de consultas en el contexto de todas las consultas que usan las tablas.

## Administración de operaciones de escritura simultáneas

Amazon Redshift permite que las tablas se puedan leer mientras se cargan o modifican de forma progresiva.

En algunas aplicaciones de inteligencia de negocios y almacenamiento de datos tradicionales, la base de datos se encuentra disponible para los usuarios solo cuando la carga nocturna se ha completado. En tales casos, no se permiten actualizaciones durante las horas de trabajo regulares, cuando se ejecutan consultas analíticas y se generan informes; no obstante, cada vez más

aplicaciones permanecen en vivo durante largos periodos del día o, incluso, todo el día, lo cual hace que la noción de una ventana de carga sea obsoleta.

Amazon Redshift es compatible con estos tipos de aplicaciones porque permite que las tablas se lean mientras se cargan o modifican de forma progresiva. Las consultas simplemente ven la última versión confirmada o una instantánea de los datos, en lugar de esperar a que se confirme la siguiente versión. Si desea que una consulta particular espere por una confirmación de otra operación de escritura, debe programarla de manera acorde.

En los temas a continuación, se describen algunos de los conceptos clave y los casos de uso que incluyen transacciones, instantáneas de base de datos, actualizaciones y comportamiento simultáneo.

Temas


- Aislamiento serializable
- Operaciones de lectura y escritura y de escritura
- Ejemplos de escritura simultánea


## Aislamiento serializable

Algunas aplicaciones no solo requieren la ejecución simultánea de consultas y cargas, sino también la capacidad de escribir en distintas tablas o en la misma tabla de manera simultánea. En este contexto, simultánea significa "superpuesta", no "programada para que se ejecute exactamente al mismo tiempo". Dos transacciones se consideran simultáneas si la segunda comienza antes de que se confirme la primera. Las operaciones concurrentes pueden originarse en sesiones diferentes controladas por el mismo usuario o por usuarios diferentes.

> **Note**
>
> Amazon Redshift admite el comportamiento de confirmación automática predeterminado en el que cada comando SQL ejecutado por separado se confirma individualmente. Si incluye un conjunto de comandos en un bloque de transacciones (definido por instrucciones BEGIN y END), el bloque se confirma como una transacción, de manera que puede revertirse de ser necesario. Las excepciones a este comportamiento son los comandos TRUNCATE y VACUUM, que confirman automáticamente todos los cambios pendientes realizados en la transacción actual.

Algunos clientes SQL emiten los comandos BEGIN y COMMIT automáticamente, por lo que el cliente controla si un grupo de instrucciones se ejecuta como una transacción o si cada instrucción individual se ejecuta como su propia transacción. Consulte la documentación de la interfaz que está utilizando. Por ejemplo, cuando se utiliza el controlador JDBC de Amazon Redshift, una cadena `PreparedStatement` de JDBC con consulta que contiene varios comandos SQL (separados por punto y coma) ejecuta todas las instrucciones como una sola transacción. Por el contrario, si utiliza SQL Workbench/J y establece AUTO COMMIT ON, si ejecuta varias instrucciones, cada una de ellas se ejecuta como una transacción propia.

Las operaciones de escritura simultáneas son compatibles con Amazon Redshift en modo protegido, con bloqueos de escritura en las tablas y el principio de aislamiento serializable. El aislamiento serializable conserva la ilusión de que una transacción que se ejecuta en una tabla es la única transacción ejecutándose en ella. Por ejemplo, dos transacciones que se ejecutan de manera simultánea, T1 y T2, deben generar los mismos resultados que al menos uno de los siguientes:

- T1 y T2 se ejecutan en serie en ese orden.
- T2 y T1 se ejecutan en serie en ese orden.

Las transacciones simultáneas son invisibles entre sí, no pueden detectar los cambios de la otra. Cada transacción simultánea creará una instantánea de base de datos al comienzo de la transacción. Una instantánea de base de datos se crea dentro de una transacción con la primera aparición de la mayoría de las instrucciones SELECT, los comandos DML como COPY, DELETE, INSERT, UPDATE y TRUNCATE, y los siguientes comandos DDL:

- ALTER TABLE (para agregar o eliminar columnas)
- CREATE TABLE
- DROP TABLE
- TRUNCATE TABLE

Si una ejecución en serie de las transacciones simultáneas produce los mismos resultados que la ejecución simultánea, estas transacciones se consideran “serializables” y pueden ejecutarse de manera segura. Si ninguna ejecución en serie de esas transacciones puede producir los mismos resultados, la transacción que ejecuta una instrucción que podría afectar la capacidad de serialización se anula y revierte.

Las tablas de catálogo del sistema (PG) y otras tablas del sistema de Amazon Redshift (STL y STV) no se bloquean en una transacción. Por lo tanto, los cambios realizados en los objetos de base de datos que surjan de las operaciones DDL y TRUNCATE son visibles con cualquier transacción simultánea cuando se confirman.

Por ejemplo, supongamos que la tabla A existe en la base de datos al momento en que comienzan dos transacciones simultáneas, T1 y T2. Supongamos que T2 devuelve una lista de tablas al seleccionar las de la tabla de catálogo PG_TABLES. Luego T1 elimina la tabla A y confirma los cambios y, a continuación, T2 vuelve a especificar las tablas. La tabla A ya no aparece en la lista. Si T2 intenta ejecutar una consulta en la tabla eliminada, Amazon Redshift devuelve un error de tipo “la relación no existe”. La consulta de catálogo que le devuelve la lista de tablas a T2 o que verifica la existencia de la tabla A no se encuentra sujeta a las mismas reglas de aislamiento que las operaciones que se realizan en las tablas de usuario.

Las transacciones de las actualizaciones de estas tablas se ejecutan en un modo de aislamiento de lectura confirmada. Las tablas de catálogo con el prefijo PG no son compatibles con el aislamiento de la instantánea.

## Aislamiento serializable para tablas de sistema y tablas de catálogo

Una instantánea de base de datos también se crea en una transacción para cualquier consulta SELECT que referencia una tabla creada por un usuario o una tabla de sistema de Amazon Redshift (STL o STV). Las consultas SELECT que no hacen referencia a ninguna tabla no crean una instantánea de base de datos de transacción nueva. Las instrucciones INSERT, DELETE y UPDATE que realizan operaciones únicamente en tablas de catálogo del sistema (PG) tampoco crean una instantánea de base de datos de transacción nueva.

## Cómo corregir errores de aislamiento serializable

ERROR:1023 DETAIL: Serializable isolation violation on a table in Redshift

Cuando Amazon Redshift detecta un error de aislamiento serializable, se muestra un mensaje de error como el siguiente.

```
ERROR:1023 DETAIL: Serializable isolation violation on table in Redshift
```

Para solucionar el error de aislamiento serializable, puede probar los métodos siguientes:

- Volver a intentar la transacción cancelada.

Amazon Redshift detectó que una carga de trabajo simultánea no es serializable. Sugiere que hay deficiencias en la lógica de la aplicación, que por lo general se pueden solucionar si se intenta realizar la transacción en la que se encontró el error de nuevo. Si el problema persiste, pruebe uno de los métodos siguientes.

- Mover cualquier operación que no tenga que estar en la misma transacción atómica fuera de la transacción.

  Este método se aplica cuando las operaciones individuales dentro de dos transacciones tienen referencias cruzadas entre sí de forma que pueden afectar al resultado de la otra transacción. Por ejemplo, las dos sesiones siguientes inician una transacción cada una.

```
Session1_Redshift=# begin;
```

```
Session2_Redshift=# begin;
```

  El resultado de la instrucción SELETCT en cada transacción se podría ver afectado por una instrucción INSERT en la otra. Dicho de otro modo, suponga que ejecuta las siguientes instrucciones en serie, en cualquier orden. En cada caso, el resultado es que una de las instrucciones SELECT devuelve una fila más que si las transacciones se ejecutaran de forma simultánea. No hay ningún orden en que las operaciones se puedan ejecutar en serie que produzca el mismo resultado que cuando se ejecutan de forma simultánea. Así, la última operación que se ejecuta da lugar a un error de aislamiento serializable.

```
Session1_Redshift=# select * from tab1;
Session1_Redshift=# insert into tab2 values (1);
```

```
Session2_Redshift=# insert into tab1 values (1);
Session2_Redshift=# select * from tab2;
```

  En muchos casos, el resultado de las instrucciones SELECT no es importante. Dicho de otro modo, la atomicidad de las operaciones en las transacciones no es importante. En estos casos, saque las instrucciones SELECT de sus transacciones, como se muestra en los ejemplos siguientes.

```
Session1_Redshift=# begin;
Session1_Redshift=# insert into tab1 values (1)
```

```
Session1_Redshift=# end;
Session1_Redshift=# select * from tab2;
```

```
Session2_Redshift # select * from tab1;
Session2_Redshift=# begin;
Session2_Redshift=# insert into tab2 values (1)
Session2_Redshift=# end;
```

En estos ejemplos, no hay ninguna referencia cruzada en las transacciones. Las dos instrucciones INSERT no se ven afectadas entre sí. En estos ejemplo, hay al menos un orden en el que las transacciones se pueden ejecutar en serie y producir el mismo resultado que si se ejecutan de manera simultánea. Estos significa que las transacciones son serializables.

- Obligue a la serialización bloqueando todas las tablas en cada sesión.

  El comando LOCK bloquea operaciones que dan lugar a errores de aislamiento serializable. Cuando utilice el comando LOCK, asegúrese de hacer lo siguiente:

  - Bloquee todas las tablas que se ven afectadas por la transacción, incluidas aquellas afectadas por instrucciones SELECT de solo lectura dentro de la transacción.
  - Bloquee las tablas en el mismo orden, con independencia del orden en que se ejecutan las operaciones.
  - Bloquee todas las tablas al principio de la transacción, antes de llevar a cabo alguna operación.
- Utilizar el aislamiento de instantáneas para transacciones simultáneas

  Utilice un comando ALTER DATABASE con aislamiento de instantáneas. Para obtener más información sobre el parámetro SNAPSHOT para ALTER DATABASE, consulte Parámetros.

### ERROR:1018 DETAIL: Relation does not exist

Cuando ejecuta operaciones simultáneas de Amazon Redshift en sesiones diferentes, se muestra un mensaje de error como el siguiente.

```
ERROR: 1018 DETAIL: Relation does not exist.
```

Las transacciones en Amazon Redshift utilizan el aislamiento de la instantánea. Una vez que se inicia una transacción, Amazon Redshift toma una instantánea de la base de datos. En todo el ciclo de vida de la transacción, la transacción realiza las operaciones en el estado de la base de datos como se refleja en la instantánea. Si la transacción realiza la lectura desde una tabla que no existe

en la instantánea, arroja el mensaje de error 1018 mostrado anteriormente. Incluso cuando otra transacción simultánea crea una tabla después de que la transacción haya tomado la instantánea, la transacción no puede realizar la lectura desde la tabla recién creada.

Para solucionar este error de serialización del aislamiento, puede intentar cambiar el inicio de la transacción a un punto en el que sepa que existe la tabla.

Si otra transacción crea la tabla, este punto será después de que se haya confirmado la transacción. Además, asegúrese de que no se haya confirmado ninguna transacción simultánea que pueda haber eliminado la tabla.

```
session1 = # BEGIN;
session1 = # DROP TABLE A;
session1 = # COMMIT;
```

```
session2 = # BEGIN;
```

```
session3 = # BEGIN;
session3 = # CREATE TABLE A (id INT);
session3 = # COMMIT;
```

```
session2 = # SELECT * FROM A;
```

La última operación que ejecuta session2 como la operación de lectura da lugar a un error de aislamiento serializable. Este error ocurre cuando session2 toma una instantánea y la tabla ya ha sido eliminada por una session1 confirmada. En otras palabras, aunque una session3 simultánea haya creado la tabla, la session2 no ve la tabla porque no está en la instantánea.

Para resolver este error, puede reordenar las sesiones de la siguiente manera.

```
session1 = # BEGIN;
session1 = # DROP TABLE A;
session1 = # COMMIT;
```

```
session3 = # BEGIN;
session3 = # CREATE TABLE A (id INT);
session3 = # COMMIT;
```

```
session2 = # BEGIN;
session2 = # SELECT * FROM A;
```

Cuando session2 toma su instantánea, session3 ya se ha confirmado y la tabla está en la base de datos. La session2 puede leer desde la tabla sin ningún error.

## Operaciones de lectura y escritura y de escritura

Puede administrar el comportamiento específico de las operaciones simultáneas de escritura al decidir cuándo y cómo ejecutar diferentes tipos de comandos. Los siguientes comandos son relevantes en este análisis:

- Comandos COPY, que realizan cargas (iniciales o incrementales).
- Comandos INSERT, que agregan una o más filas por vez.
- Comandos UPDATE, que modifican filas existentes.
- Comandos DELETE, que eliminan filas.

Las operaciones COPY e INSERT son operaciones de escritura pura. Las operaciones DELETE y UPDATE son operaciones de lectura/escritura (para eliminar o actualizar filas, primero hay que leerlas). Los resultados de las operaciones simultáneas de escritura dependen de los comandos específicos que se están ejecutando simultáneamente.

Las operaciones UPDATE y DELETE se comportan de manera diferente porque dependen de una lectura inicial de la tabla antes de cualquier escritura. Dado que las transacciones simultáneas son invisibles entre sí, tanto UPDATE como DELETE deben leer una snapshot de los datos de la última confirmación. Cuando la primera operación UPDATE o DELETE levanta el bloqueo, la segunda debe determinar si los datos con los que trabajará son potencialmente obsoletos. No lo serán, porque la segunda transacción no obtiene su snapshot de los datos sino hasta después de que la primera transacción haya levantado su bloqueo.

### Posible situación de bloqueo para transacciones de escritura simultáneas que involucran múltiples tablas

Cuando las transacciones involucren actualizaciones de más de una tabla, existe la posibilidad de que transacciones que se ejecutan de manera simultánea se bloqueen al querer escribir en el mismo conjunto de tablas. Una transacción levanta todos los bloqueos de tabla a la vez, ya sea cuando confirma o cuando se revierte; no los libera de a uno.

Por ejemplo, supongamos que las transacciones T1 y T2 comienzan, aproximadamente, al mismo tiempo. Si T1 comienza a escribir en la tabla A y T2 comienza a escribir en la tabla B, ambas transacciones pueden continuar sin conflictos. No obstante, si T1 finaliza la escritura de la tabla A y necesita comenzar a escribir en la tabla B, no podrá hacerlo porque T2 aún la bloquea. Del mismo modo, si T2 termina de escribir en la tabla B y necesita comenzar a escribir en la tabla A, tampoco podrá hacerlo porque T1 aún la bloquea. Como ninguna de las transacciones puede liberar los bloqueos sino hasta que todas las operaciones de escritura se hayan confirmado, ninguna transacción puede avanzar. Para evitar este tipo de bloqueo, debe programar las operaciones de escritura simultáneas con sumo cuidado. Por ejemplo, siempre debe actualizar las tablas en el mismo orden en las transacciones y, si especifica bloqueos, bloquee las tablas en el mismo orden antes de realizar cualquier operación DML.

## Posible situación de bloqueo para transacciones de escritura simultáneas que involucran una sola tabla

En un entorno de aislamiento de instantáneas, pueden producirse bloqueos al ejecutar transacciones de escritura simultáneas en la misma tabla. El bloqueo de aislamiento de instantáneas ocurre cuando las instrucciones INSERT o COPY simultáneas comparten un bloqueo y avanzan, y otra instrucción necesita realizar una operación (UPDATE, DELETE, MERGE o DDL) que requiere un bloqueo exclusivo en la misma tabla.

Veamos la siguiente situación:

Transacción 1 (T1):

```
INSERT/COPY INTO table_A;
```

Transacción 2 (T2):

```
INSERT/COPY INTO table_A;
            <UPDATE/DELETE/MERGE/DDL statement> table_A
```

Puede producirse un bloqueo cuando varias transacciones con operaciones INSERT o COPY se ejecutan simultáneamente en la misma tabla con un bloqueo compartido, y una de esas transacciones sigue la operación de escritura pura con una operación que requiere un bloqueo exclusivo, como una instrucción UPDATE, MERGE, DELETE o DDL.

Para evitar el bloqueo en estas situaciones, puede separar las instrucciones que requieren un bloqueo exclusivo (instrucciones UPDATE/MERGE/DELETE/DDL) en una transacción diferente

para que las instrucciones INSERT/COPY puedan progresar simultáneamente y las instrucciones que requieren bloqueos exclusivos puedan ejecutarse después de ellas. De forma alternativa, para transacciones con instrucciones INSERT/COPY y MERGE/UPDATE/MERGE en la misma tabla, puede incluir lógica de reintento en las aplicaciones para evitar posibles bloqueos.

## Ejemplos de escritura simultánea

En los siguientes ejemplos de pseudocódigo, se demuestra cómo, cuando se ejecutan de manera simultánea, las transacciones avanzan o esperan.

### Ejemplos de escritura simultánea con aislamiento serializable

Operaciones COPY simultáneas en la misma tabla con aislamiento serializable

La transacción 1 copia filas en la tabla LISTING:

```
begin;
copy listing from ...;
end;
```

La transacción 2 comienza de manera simultánea en una sesión por separado y copia más filas en la tabla LISTING. La transacción 2 debe esperar hasta que la transacción 1 levante el bloqueo de escritura de la tabla LISTING; luego, podrá avanzar.

```
begin;
[waits]
copy listing from ;
end;
```

El mismo comportamiento se observaría si una o ambas transacciones tuvieran un comando INSERT, en lugar de COPY.

Operaciones DELETE simultáneas desde la misma tabla con aislamiento serializable

La transacción 1 elimina filas de una tabla:

```
begin;
delete from listing where ...;
end;
```

La transacción 2 comienza de manera simultánea y elimina filas de la misma tabla. Tendrá éxito porque espera a que la transacción 1 termine antes de probar eliminar filas.

```
begin
[waits]
delete from listing where ;
end;
```

El mismo comportamiento se observaría si una o ambas transacciones tuvieran un comando UPDATE para la misma tabla, en lugar de DELETE.

Transacciones simultáneas con una mezcla de operaciones de lectura y escritura con aislamiento serializable

En este ejemplo, la transacción 1 elimina filas de la tabla USERS, vuelve a cargar la tabla, ejecuta una consulta COUNT(*) y, luego, ANALYZE, antes de confirmar:

```
begin;
delete one row from USERS table;
copy ;
select count(*) from users;
analyze ;
end;
```

Mientras tanto, comienza la transacción 2. Esta transacción prueba copiar filas adicionales en la tabla USERS, analizarla y, luego, ejecutar la misma consulta COUNT(*) que la primera transacción:

```
begin;
[waits]
copy users from ...;
select count(*) from users;
analyze;
end;
```

La segunda transacción tendrá éxito porque debe esperar a que la primera termine. Su consulta COUNT devolverá el recuento en función de la carga que ha completado.

## Ejemplos de escritura simultánea con aislamiento de instantánea

Operaciones de copia simultáneas en la misma tabla con aislamiento de instantánea

La transacción 1 copia filas en la tabla LISTING:

```
begin;
copy listing from ...;
end;
```

La transacción 2 comienza de manera simultánea en una sesión por separado y copia más filas en la tabla LISTING. La transacción 2 puede progresar simultáneamente hasta que cualquiera de las transacciones necesite escribir datos en la tabla de destino `listing`, momento en el que se ejecutarán secuencialmente.

```
begin;
//When the COPY statement from T1 needs to write data to the table, the COPY statement
 from T2 waits.
copy listing from ...;
end;
```

El mismo comportamiento se observaría si una o ambas transacciones tuvieran un comando INSERT, en lugar de COPY.

Operaciones DELETE simultáneas desde la misma tabla con aislamiento de instantánea

Las operaciones DELETE y UPDATE simultáneas desde la misma tabla con aislamiento de instantánea se ejecutan igual que las operaciones ejecutadas con aislamiento serializable.

Transacciones simultáneas con una mezcla de operaciones de lectura y escritura con aislamiento de instantánea

Las transacciones simultáneas que se ejecutan con mezclas de operaciones con aislamiento de instantánea se ejecutan igual que las transacciones con mezclas de operaciones que se ejecutan con aislamiento serializable.

# Tutorial: Carga de datos desde Amazon S3

En este tutorial, siga el proceso de creación de datos en las tablas de bases de datos de Amazon Redshift desde archivos de datos en un bucket de Amazon S3 de principio a fin.

En este tutorial, aprenderá a hacer lo siguiente:

- Descargar archivos de datos que utilicen formatos de ancho fijo, con caracteres delimitados y de valores separados por comas (CSV).

- crear un bucket de Amazon S3 y, luego, cargar los archivos de datos en él
- lanzar un clúster de Amazon Redshift y crear tablas de base de datos
- usar los comandos COPY para cargar las tablas desde los archivos de datos en Amazon S3
- Solucionar errores de carga y modificar los comandos COPY para corregir los errores.

## Requisitos previos

Necesita los siguientes requisitos previos:

- Una cuenta de AWS para lanzar un clúster de Amazon Redshift y crear un bucket en Amazon S3.
- Sus credenciales de AWS (rol de IAM) para cargar datos de prueba de Amazon S3. Si necesita un nuevo rol de IAM, diríjase a Creación de roles de IAM.
- Un cliente SQL como el editor de consultas de la consola de Amazon Redshift.

Este tutorial está diseñado para que pueda realizarlo por sí mismo. Además de este tutorial, recomendamos completar los siguientes tutoriales para obtener un conocimiento más completo sobre cómo diseñar y utilizar las bases de datos de Amazon Redshift:

- En la Guía de introducción a Amazon Redshift, se detalla el proceso de la creación de un clúster de Amazon Redshift y de la carga de datos de muestra.

## Descripción general

Puede agregar datos a las tablas de Amazon Redshift mediante el uso de un comando INSERT o uno COPY. En la escala y la velocidad de un almacenamiento de datos de Amazon Redshift, el comando COPY es muchas veces más rápido y eficiente que los comandos INSERT.

El comando COPY utiliza la arquitectura de procesamiento masivo en paralelo (MPP) de Amazon Redshift para leer y cargar datos en paralelo desde varios orígenes de datos. Puede cargar datos desde archivos de datos en Amazon S3, Amazon EMR o cualquier alojamiento remoto que sea accesible a través de una conexión Secure Shell (SSH). También puede cargarlos directamente desde una tabla de Amazon DynamoDB.

En este tutorial, se utiliza el comando COPY para cargar datos desde Amazon S3. Muchos de los principios presentados aquí también aplican para la carga desde otros orígenes de datos.

Para obtener más información acerca del uso del comando COPY, consulte los siguientes recursos:

- Prácticas recomendadas de Amazon Redshift para la carga de datos
- Carga de datos desde Amazon EMR
- Carga de datos desde hosts remotos
- Carga de datos desde una tabla de Amazon DynamoDB

## Paso 1: creación de un clúster

Puede omitir este paso si ya dispone del clúster que desea utilizar.

Para los ejercicios de este tutorial, use un clúster de cuatro nodos.

Para crear un clúster

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.

   En el menú de navegación, elija Panel de clústeres aprovisionados.

   > **Important**
   >
   > Asegúrese de que el usuario cuente con los permisos necesarios para realizar las operaciones del clúster. Para obtener información sobre la concesión de los permisos necesarios, consulte Autorización a Amazon Redshift para obtener acceso a los servicios de AWS.

2. En la parte superior derecha, seleccione la región de AWS en la que quiera crear el clúster. Para realizar este tutorial, elija EE. UU. Oeste (Oregón).
3. En el menú de navegación, elija Clusters (Clústeres) y, a continuación, elija Create cluster (Crear clúster). Se abrirá la página Create cluster (Crear clúster).
4. En la página Create cluster (Crear un clúster), ingrese los parámetros para su clúster. Elija sus propios valores para los parámetros, excepto cambiar los siguientes valores:

   - Elija **`dc2.large`** para el tipo de nodo.
   - Elija **`4`** para la opción Number of nodes (Número de nodos).
   - En la sección Cluster permissions (Permisos del clúster), seleccione un rol de IAM de los Available IAM roles (Roles de IAM disponibles). Este rol debería ser uno que haya creado previamente y que tenga acceso a Amazon S3. Luego, elija Associate IAM role (Asociar un

rol de IAM) para agregarlo a la lista de Associated IAM roles (Roles de IAM asociados) del clúster.

5. Elija Create cluster.

Siga los pasos de la Guía de introducción a Amazon Redshift para conectarse al clúster desde un cliente SQL y probar una conexión. No necesita completar los pasos de introducción restantes para crear tablas, cargar datos y probar consultas de ejemplo.

## Paso 2: Descargar los archivos de datos

En este paso, se descarga un conjunto de archivos de datos de muestra en su equipo. En el paso siguiente, se cargan los archivos a un bucket de Amazon S3.

Descarga de los archivos de datos

1. Descargue el archivo comprimido: LoadingDataSampleFiles.zip.
2. Extraiga los archivos en una carpeta de su equipo.
3. Verifique que la carpeta contiene los siguientes archivos.

```
customer-fw-manifest
customer-fw.tbl-000
customer-fw.tbl-000.bak
customer-fw.tbl-001
customer-fw.tbl-002
customer-fw.tbl-003
customer-fw.tbl-004
customer-fw.tbl-005
customer-fw.tbl-006
customer-fw.tbl-007
customer-fw.tbl.log
dwdate-tab.tbl-000
dwdate-tab.tbl-001
dwdate-tab.tbl-002
dwdate-tab.tbl-003
dwdate-tab.tbl-004
dwdate-tab.tbl-005
dwdate-tab.tbl-006
dwdate-tab.tbl-007
part-csv.tbl-000
part-csv.tbl-001
```

```
part-csv.tbl-002
part-csv.tbl-003
part-csv.tbl-004
part-csv.tbl-005
part-csv.tbl-006
part-csv.tbl-007
```

## Paso 3: Cargar los archivos en un bucket de Amazon S3

En este paso, cree un bucket de Amazon S3 y cargue los archivos de datos en él.

Para cargar los archivos en un bucket de Amazon S3

1. Cree un bucket de Amazon S3.

   Para obtener más información acerca de la creación de un bucket, consulte Crear un bucket en la Guía del usuario de Amazon Simple Storage Service.

   a. Inicie sesión en la AWS Management Console y abra la consola de Amazon S3 en https://console.aws.amazon.com/s3/.

   b. Elija Crear bucket.

   c. Elija un valor para Región de AWS.

      Cree el bucket en la misma región del clúster. Si el clúster se encuentra en la región Oeste de EE. UU. (Oregón), elija US West (Oregon) Region (us-west-2) (Región Oeste de EE. UU. [Oregón] [us-west-2]).

   d. En el cuadro Nombre del bucket del cuadro de diálogo Crear bucket, escriba el nombre del bucket.

      El nombre de bucket que elija debe ser único entre todos los nombres de buckets existentes en Amazon S3. Una forma de garantizar su unicidad es agregar el nombre de su organización delante del nombre del bucket. Los nombres de los buckets deben cumplir una serie de reglas. Para obtener más información, consulte Restricciones y limitaciones de buckets en la Guía del usuario de Amazon Simple Storage Service.

   e. Elija los valores predeterminados recomendados para el resto de las opciones.

   f. Elija Crear bucket.

Cuando Amazon S3 cree con éxito el bucket, la consola mostrará el bucket vacío en el panel Buckets.

2. Crear una carpeta.

   a. Seleccione el nombre del nuevo bucket.

   b. Elija el botón Crear carpeta.

   c. Asigne un nombre para la nueva carpeta **`load`**.

   > **Note**
   >
   > El bucket que ha creado no se encuentra en un entorno de pruebas. En este ejercicio, se agregan objetos a un bucket real. Se le cobra una cantidad nominal según el tiempo durante el cuál almacena los objetos en el bucket. Para obtener más información acerca de los precios de Amazon S3, visite la página Precios de Amazon S3.

3. Cargue los archivos de datos en el bucket de Amazon S3 nuevo.

   a. Elija el nombre de la carpeta de datos.

   b. En el asistente Cargar, elija Agregar archivos.

      Siga las instrucciones de la consola de Amazon S3 para cargar todos los archivos que descargó y extrajo,

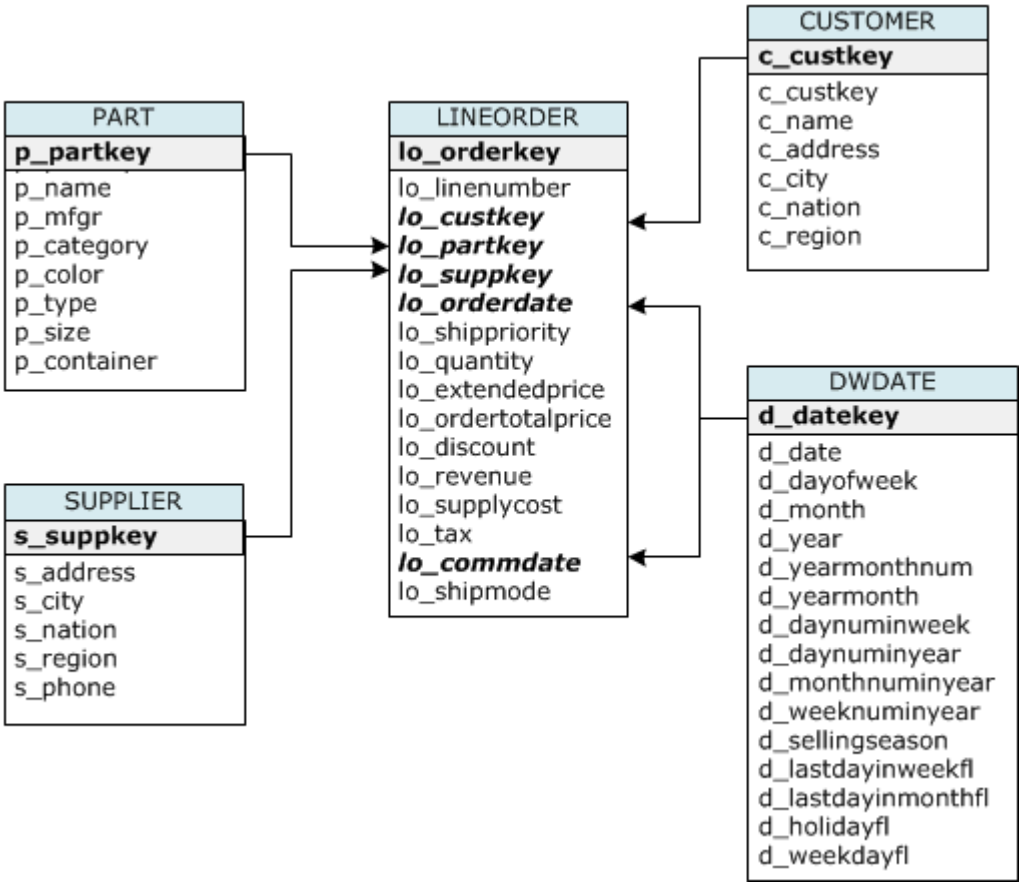
   c. Seleccione Cargar.

de usuario autentic.

El comando COPY de Amazon Redshift debe tener acceso para leer los objetos de los archivos del bucket de Amazon S3. Si utiliza las mismas credenciales de usuario para crear el bucket de Amazon S3 y ejecutar el comando COPY de Amazon Redshift, el comando COPY tiene todos los permisos necesarios. Si desea utilizar credenciales de usuario diferentes, puede conceder acceso mediante los controles de acceso de Amazon S3. El comando COPY de Amazon Redshift requiere como mínimo los permisos de ListBucket y GetObject para obtener acceso a los objetos de los archivos del bucket de Amazon S3. Para obtener más información acerca del control del acceso a los recursos de Amazon S3, visite Administración de permisos de acceso para los recursos de Amazon S3.

# Paso 4: Crear las tablas de muestra

En este tutorial, se utiliza un conjunto de tablas basadas en el esquema de Star Schema Benchmark (SSB). En el siguiente diagrama se muestra el modelo de datos en SSB.




Puede que las tablas de SSB ya existan en la base de datos actual. Si es el caso, deberá eliminarlas para quitarlas de la base de datos antes de crearlas con los comandos CREATE TABLE en el siguiente paso. Las tablas utilizadas en este tutorial pueden tener atributos diferentes a las tablas existentes.

Creación de las tablas de muestra

1. Para eliminar las tablas SSB, ejecute los siguientes comandos en su cliente SQL.

```
drop table part cascade;
drop table supplier;
drop table customer;
drop table dwdate;
drop table lineorder;
```

2. Ejecute los siguientes comandos CREATE TABLE en el cliente SQL.

```
CREATE TABLE part
(
  p_partkey     INTEGER NOT NULL,
  p_name        VARCHAR(22) NOT NULL,
  p_mfgr        VARCHAR(6),
  p_category    VARCHAR(7) NOT NULL,
  p_brand1      VARCHAR(9) NOT NULL,
  p_color       VARCHAR(11) NOT NULL,
  p_type        VARCHAR(25) NOT NULL,
  p_size        INTEGER NOT NULL,
  p_container   VARCHAR(10) NOT NULL
);

CREATE TABLE supplier
(
  s_suppkey   INTEGER NOT NULL,
  s_name      VARCHAR(25) NOT NULL,
  s_address   VARCHAR(25) NOT NULL,
  s_city      VARCHAR(10) NOT NULL,
  s_nation    VARCHAR(15) NOT NULL,
  s_region    VARCHAR(12) NOT NULL,
  s_phone     VARCHAR(15) NOT NULL
);

CREATE TABLE customer
(
  c_custkey      INTEGER NOT NULL,
  c_name         VARCHAR(25) NOT NULL,
  c_address      VARCHAR(25) NOT NULL,
  c_city         VARCHAR(10) NOT NULL,
  c_nation       VARCHAR(15) NOT NULL,
  c_region       VARCHAR(12) NOT NULL,
  c_phone        VARCHAR(15) NOT NULL,
  c_mktsegment   VARCHAR(10) NOT NULL
);

CREATE TABLE dwdate
(
  d_datekey            INTEGER NOT NULL,
  d_date               VARCHAR(19) NOT NULL,
  d_dayofweek          VARCHAR(10) NOT NULL,
```

```
  d_month              VARCHAR(10) NOT NULL,
  d_year               INTEGER NOT NULL,
  d_yearmonthnum       INTEGER NOT NULL,
  d_yearmonth          VARCHAR(8) NOT NULL,
  d_daynuminweek       INTEGER NOT NULL,
  d_daynuminmonth      INTEGER NOT NULL,
  d_daynuminyear       INTEGER NOT NULL,
  d_monthnuminyear     INTEGER NOT NULL,
  d_weeknuminyear      INTEGER NOT NULL,
  d_sellingseason      VARCHAR(13) NOT NULL,
  d_lastdayinweekfl    VARCHAR(1) NOT NULL,
  d_lastdayinmonthfl   VARCHAR(1) NOT NULL,
  d_holidayfl          VARCHAR(1) NOT NULL,
  d_weekdayfl          VARCHAR(1) NOT NULL
);
CREATE TABLE lineorder
(
  lo_orderkey          INTEGER NOT NULL,
  lo_linenumber        INTEGER NOT NULL,
  lo_custkey           INTEGER NOT NULL,
  lo_partkey           INTEGER NOT NULL,
  lo_suppkey           INTEGER NOT NULL,
  lo_orderdate         INTEGER NOT NULL,
  lo_orderpriority     VARCHAR(15) NOT NULL,
  lo_shippriority      VARCHAR(1) NOT NULL,
  lo_quantity          INTEGER NOT NULL,
  lo_extendedprice     INTEGER NOT NULL,
  lo_ordertotalprice   INTEGER NOT NULL,
  lo_discount          INTEGER NOT NULL,
  lo_revenue           INTEGER NOT NULL,
  lo_supplycost        INTEGER NOT NULL,
  lo_tax               INTEGER NOT NULL,
  lo_commitdate        INTEGER NOT NULL,
  lo_shipmode          VARCHAR(10) NOT NULL
);
```

## Paso 5: Ejecutar los comandos COPY

Ejecute los comandos COPY para cargar cada una de las tablas en el esquema SSB. Los ejemplos del comando COPY muestran cómo cargar desde diferentes formatos de archivo, utilizar varias opciones de comando COPY y solucionar los errores de carga.

## Sintaxis del comando COPY

La sintaxis básica del comando COPY es la siguiente.

```
COPY table_name [ column_list ] FROM data_source CREDENTIALS access_credentials
 [options]
```

Para ejecutar un comando COPY, proporcione los siguientes valores.

Nombre de la tabla

La tabla de destino para el comando COPY. La tabla ya debe existir en la base de datos. La tabla puede ser temporal o persistente. El comando COPY adjunta los nuevos datos de entrada a cualquier fila existente en la tabla.

Lista de columnas

De forma predeterminada, COPY carga los campos desde los datos de origen a las columnas de la tabla en orden. Si lo desea, también puede especificar una lista de columnas, que es una lista de nombres de columnas separadas por una coma, para asignar los campos de datos a columnas específicas. No utiliza listas de columnas en este tutorial. Para obtener más información, consulte Column List en COPY command reference.

Origen de datos

Puede utilizar el comando COPY para cargar datos desde un bucket de Amazon S3, un clúster de Amazon EMR, un alojamiento remoto que utilice una conexión SSH o una tabla de Amazon DynamoDB. Para este tutorial, realice la carga a partir de archivos de datos en un bucket de Amazon S3. Cuando realice cargas desde Amazon S3, debe proporcionar el nombre del bucket y la ubicación de los archivos de datos. Para ello, proporcione o bien una ruta de objecto para los archivos de datos o bien la ubicación de un archivo de manifiesto que enumere explícitamente cada archivo de datos y su ubicación.

- Prefijo de clave

  Un objeto almacenado en Amazon S3 se identifica de manera única por una clave de objeto, que incluye el nombre del bucket, de las carpetas, si hay alguna, y de los objetos. El prefijo de clave hace referencia a un conjunto de objetos con el mismo prefijo. La ruta de objeto es un prefijo de clave que el comando COPY utiliza para cargar todos los objetos que comparten el prefijo de clave. Por ejemplo, el prefijo de clave `custdata.txt` puede referirse a un archivo

individual o a un conjunto de archivos, incluidos `custdata.txt.001`, `custdata.txt.002` y así sucesivamente.

- Archivo de manifiesto

  En algunos casos, es posible que necesite cargar archivos con diferentes prefijos; por ejemplo, de múltiples buckets o carpetas. En otros, puede que necesite excluir archivos con un mismo prefijo. En estos casos, puede utilizar un archivo de manifiesto. El archivo de manifiesto muestra explícitamente cada archivo de carga y su clave de objeto única. Se utiliza un archivo de manifiesto para cargar la tabla PART más adelante en este tutorial.

Credenciales

Para obtener acceso a los recursos de AWS que contienen los datos que se van a cargar, debe proporcionar las credenciales de acceso de AWS de un usuario con privilegios suficientes. Estas credenciales incluyen el Nombre de recurso de Amazon (ARN) de un rol de IAM. Para cargar datos desde Amazon S3, las credenciales deben incluir los permisos de ListBucket y GetObject. Se requerirán credenciales adicionales si los datos están cifrados. Para obtener más información, consulte Parámetros de autorización en COPY command reference. Para obtener más información acerca de la administración del acceso, visite Administración de permisos de acceso a los recursos de Amazon S3.

Opciones

Puede especificar un número de parámetros con el comando COPY para especificar los formatos de archivo, administrar los formatos de datos, administrar errores y controlar otras características. En este tutorial, utilice las siguientes opciones y características del comando COPY:

- Prefijo de clave

  Para obtener información sobre cómo cargar desde varios archivos especificando un prefijo de clave, consulte Carga de tabla PART con NULL AS.

- Formato CSV

  Para obtener información sobre cómo cargar datos en formato CSV, consulte Carga de tabla PART con NULL AS.

- NULL AS

  Para obtener información sobre cómo cargar PART mediante la opción NULL AS, consulte Carga de tabla PART con NULL AS.

- Formato delimitado por caracteres

  Para obtener información sobre cómo utilizar la opción DELIMITER, consulte Las opciones DELIMITER y REGION.

- REGION

  Para obtener información sobre cómo utilizar la opción REGION, consulte Las opciones DELIMITER y REGION.

- Ancho de formato fijo

  Para obtener información sobre cómo cargar la tabla CUSTOMER a partir de datos de ancho fijo, consulte Carga de tabla CUSTOMER con MANIFEST.

- MAXERROR

  Para obtener información sobre cómo utilizar la opción MAXERROR, consulte Carga de tabla CUSTOMER con MANIFEST.

- ACCEPTINVCHARS

  Para obtener información sobre cómo utilizar la opción ACCEPTINVCHARS, consulte Carga de tabla CUSTOMER con MANIFEST.

- MANIFEST

  Para obtener información sobre cómo utilizar la opción MANIFEST, consulte Carga de tabla CUSTOMER con MANIFEST.

- DATEFORMAT

  Para obtener información sobre cómo utilizar la opción DATEFORMAT, consulte Carga de la tabla DWDATE con DATEFORMAT.

- GZIP, LZOP y BZIP2

  Para obtener información sobre cómo comprimir los archivos, consulte Carga de varios archivos de datos.

- COMPUPDATE

  Para obtener información sobre cómo utilizar la opción COMPUPDATE, consulte Carga de varios archivos de datos.

- Múltiples archivos

Para obtener información sobre cómo cargar varios archivos, consulte Carga de varios archivos de datos.

## Carga de las tablas SSB

Utilice los siguientes comandos COPY para cargar cada una de las tablas en el esquema SSB. El comando de cada tabla muestra diferentes técnicas de solución de problemas y opciones de COPY.

Siga estos pasos para cargar las tablas SSB:

1. Sustituya el nombre del bucket y las credenciales de AWS.
2. Carga de tabla PART con NULL AS
3. Carga de tabla CUSTOMER con MANIFEST
4. Carga de la tabla DWDATE con DATEFORMAT

Sustituya el nombre del bucket y las credenciales de AWS.

En este tutorial, los comandos COPY se presentan en el siguiente formato.

```
copy table from 's3://<your-bucket-name>/load/key_prefix'
credentials 'aws_iam_role=arn:aws:iam::<aws-account-id>:role/<role-name>'
options;
```

Para cada comando COPY, haga lo siguiente:

1. Sustituya *<your-bucket-name>* por el nombre de un bucket de la misma región que el clúster.

   En este paso se asume que el bucket y el clúster están en la misma región. También puede especificar la región utilizando la opción REGION con el comando COPY.
2. Reemplace *<aws-account-id>* y *<role-name>* por Cuenta de AWS y con el rol de IAM. El segmento de la cadena de credenciales que está delimitado entre comillas simples no debe contener espacios ni saltos de línea. Tenga en cuenta que el formato del ARN puede diferir ligeramente del de la muestra. Es mejor copiar el ARN del rol desde la consola de IAM para asegurarse de que es correcto cuando ejecute los comandos COPY.

### Carga de tabla PART con NULL AS

En este paso, utilice las opciones CSV y NULL AS para cargar la tabla PART.

El comando COPY puede cargar datos desde varios archivos en paralelo, lo que es mucho más rápido que cargar desde un archivo individual. Para demostrar este principio, los datos de cada tabla en este tutorial se dividen en ocho archivos, aunque los archivos sean muy pequeños. En un paso posterior, se compara la diferencia de tiempo entre cargar desde un archivo individual y cargar desde varios archivos. Para obtener más información, consulte Dividir los datos de carga.

#### Prefijo de clave

Puede cargar desde varios archivos especificando un prefijo de clave para el conjunto de archivos o enumerando explícitamente los archivos en un archivo de manifiesto. En este paso, se utiliza un prefijo de clave. En un paso posterior, se utiliza un archivo de manifiesto. El prefijo de clave `'s3://amzn-s3-demo-bucket/load/part-csv.tbl'` carga el siguiente conjunto de los archivos en la carpeta `load`.

```
part-csv.tbl-000
part-csv.tbl-001
part-csv.tbl-002
part-csv.tbl-003
part-csv.tbl-004
part-csv.tbl-005
part-csv.tbl-006
part-csv.tbl-007
```

#### Formato CSV

CSV, que significa valores separados por comas, es un formato común utilizado para importar y exportar datos de hojas de cálculo. CSV es más flexible que el formato delimitado por comas porque le permite incluir cadenas entre comillas dentro de los campos. El carácter de comilla predeterminado para COPY desde un formato CSV es una comilla recta doble ( " ), pero puede especificar otro con la opción QUOTE AS. Cuando utilice el carácter de comilla dentro del campo, encierre el carácter con otro carácter de comilla.

El siguiente fragmento de una archivo de datos con formato CSV para la tabla PART muestra cadenas encerradas entre comillas dobles (`"LARGE ANODIZED BRASS"`). También muestra un string comprendido entre dos comillas dobles dentro de un string encomillado (`"MEDIUM ""BURNISHED"" TIN"`).

```
15,dark sky,MFGR#3,MFGR#47,MFGR#3438,indigo,"LARGE ANODIZED BRASS",45,LG CASE
22,floral beige,MFGR#4,MFGR#44,MFGR#4421,medium,"PROMO, POLISHED BRASS",19,LG DRUM
23,bisque slate,MFGR#4,MFGR#41,MFGR#4137,firebrick,"MEDIUM ""BURNISHED"" TIN",42,JUMBO
 JAR
```

Los datos de la tabla PART contienen caracteres que pueden provocar que COPY falle. En este ejercicio, se identifican los errores y se corrigen.

Para cargar datos que estén en formato CSV, agregue `csv` al comando COPY. Ejecute el siguiente comando para cargar la tabla PART.

```
copy part from 's3://<your-bucket-name>/load/part-csv.tbl'
credentials 'aws_iam_role=arn:aws:iam::<aws-account-id>:role/<role-name>'
csv;
```

Podría obtener un mensaje de error similar al siguiente.

```
An error occurred when executing the SQL command:
copy part from 's3://amzn-s3-demo-bucket/load/part-csv.tbl'
credentials' ...

ERROR: Load into table 'part' failed.  Check 'stl_load_errors' system table for
 details. [SQL State=XX000]

Execution time: 1.46s

1 statement(s) failed.
1 statement(s) failed.
```

Para obtener más información acerca del error, consulte la tabla STL_LOAD_ERRORS. La siguiente consulta utiliza la función SUBSTRING para acortar columnas y facilitar su legibilidad y utiliza LIMIT 10 para reducir el número de filas que se devuelven. Puede ajustar los valores en `substring(filename,22,25)` para permitir la longitud del nombre de bucket.

```
select query, substring(filename,22,25) as filename,line_number as line,
substring(colname,0,12) as column, type, position as pos, substring(raw_line,0,30) as
 line_text,
substring(raw_field_value,0,15) as field_text,
substring(err_reason,0,45) as reason
from stl_load_errors
```

```
order by query desc
limit 10;
```

```
 query  |    filename      | line |  column   |    type    | pos |
--------+------------------------+----------+------------+------------+-----+----
 333765 | part-csv.tbl-000 |    1 |           |            |   0 |

 line_text        | field_text |                    reason
------------------+------------+---------------------------------------------
 15,NUL next,     |            | Missing newline: Unexpected character 0x2c f
```

NULL AS

Los archivos de datos `part-csv.tbl` utilizan el carácter de terminación NUL (`\x000` o `\x0`) para indicar los valores NULL.

> **Note**
>
> A pesar de tener nombres muy similares, NUL y NULL no son lo mismo. NUL es un carácter UTF-8 con punto de código `x000` que suele utilizarse para indicar el fin de registro (EOR). NULL es un valor SQL que representa una ausencia de datos.

De manera predeterminada COPY trata el carácter de terminación NUL como un carácter EOR y termina el registro, lo que suele ocasionar resultados inesperados o un error. No hay un método estándar único para indicar NULL en datos de texto. Por lo tanto, la opción de comando NULL AS COPY le permite especificar qué carácter sustituir por NULL cuando carga la tabla. En este ejemplo, desea que COPY trate el carácter de terminación NUL como un valor NULL.

> **Note**
>
> La columna de la tabla que recibe el valor NULL debe estar configurada como nullable. Es decir, no debe incluir la restricción NOT NULL en la especificación CREATE TABLE.

Para cargar PART por medio de la opción NULL AS, ejecute el siguiente comando COPY.

```
copy part from 's3://<your-bucket-name>/load/part-csv.tbl'
credentials 'aws_iam_role=arn:aws:iam::<aws-account-id>:role/<role-name>'
```

```
csv
null as '\000';
```

Para verificar que COPY cargó los valores NULL, ejecute el siguiente comando para seleccionar solo las filas que contienen NULL.

```
select p_partkey, p_name, p_mfgr, p_category from part where p_mfgr is null;
```

```
 p_partkey |  p_name  | p_mfgr | p_category
-----------+----------+--------+------------
        15 | NUL next |        | MFGR#47
        81 | NUL next |        | MFGR#23
       133 | NUL next |        | MFGR#44
(2 rows)
```

Las opciones DELIMITER y REGION

Las opciones DELIMITER y REGION son importantes para entender cómo se cargan los datos.

Formato delimitado por caracteres

Los campos en un archivo delimitado por caracteres están separados por un carácter específico, como un carácter de barra vertical ( | ), una coma ( , ) o una pestaña ( \t ). Los archivos delimitados por caracteres pueden utilizar cualquier carácter único ASCII, incluido uno de los caracteres ASCII no imprimibles, como el delimitador. El carácter delimitador se especifica mediante la opción DELIMITER. El delimitador predeterminado es el carácter de barra vertical ( | ).

El siguiente fragmento de los datos para la tabla SUPPLIER utiliza el formato delimitado por la barra vertical.

```
1|1|257368|465569|41365|19950218|2-HIGH|0|17|2608718|9783671|4|2504369|92072|2|
19950331|TRUCK
1|2|257368|201928|8146|19950218|2-HIGH|0|36|6587676|9783671|9|5994785|109794|6|
19950416|MAIL
```

REGION

Siempre que sea posible, debe ubicar los datos de carga en la misma región de AWS que el clúster de Amazon Redshift. Si los datos y el clúster se encuentran en la misma región, se reduce la latencia

y se evitan los costos de la transferencia de datos entre regiones. Para obtener más información, consulte Prácticas recomendadas de Amazon Redshift para la carga de datos.

Si debe cargar datos de una región de AWS diferente, utilice la opción REGION para especificar la región de AWS en la que se ubican los datos de carga. Si especifica una región, todos los datos de carga, incluidos los archivos de manifiesto, deben encontrarse en la región mencionada. Para obtener más información, consulte REGION.

Por ejemplo, si el clúster se encuentra en la región Este de EE. UU. (Norte de Virginia) y su bucket de Amazon S3 se encuentra en la región Oeste de EE. UU. (Oregón), el siguiente comando COPY muestra cómo cargar la tabla SUPPLIER a partir de datos delimitados por la barra vertical.

```
copy supplier from 's3://amzn-s3-demo-bucket/ssb/supplier.tbl'
credentials 'aws_iam_role=arn:aws:iam::<aws-account-id>:role/<role-name>'
delimiter '|'
gzip
region 'us-west-2';
```

Carga de tabla CUSTOMER con MANIFEST

En este paso, se utilizan las opciones FIXEDWIDTH, MAXERROR, ACCEPTINVCHARS Y MANIFEST para cargar la tabla CUSTOMER.

Los datos de muestra para este ejercicio contienen caracteres que provocan errores cuando COPY trata de cargarlos. Utilice la opción MAXERRORS y la tabla de sistema STL_LOAD_ERRORS para solucionar los errores de carga y luego utilice las opciones ACCEPTINVCHARS y MANIFEST para eliminar los errores.

Formato de ancho fijo

El formato de ancho fijo define cada campo como un número fijo de caracteres, en lugar de separar los campos con un delimitador. El siguiente fragmento de los datos para la tabla CUSTOMER utiliza el formato de ancho fijo.

```
1   Customer#000000001   IVhzIApeRb           MOROCCO  0MOROCCO  AFRICA      25-705
2   Customer#000000002   XSTf4,NCwDVaWNe6tE   JORDAN   6JORDAN   MIDDLE EAST 23-453
3   Customer#000000003   MG9kdTD              ARGENTINA5ARGENTINAAMERICA     11-783
```

El orden de los pares etiqueta/ancho debe coincidir con el orden de las columnas de la tabla de forma exacta. Para obtener más información, consulte FIXEDWIDTH.

La cadena de especificaciones de ancho fijo para la tabla CUSTOMER es la siguiente.

```
fixedwidth 'c_custkey:10, c_name:25, c_address:25, c_city:10, c_nation:15,
c_region :12, c_phone:15,c_mktsegment:10'
```

Para cargar la tabla CUSTOMER desde los datos de ancho fijo, ejecute el siguiente comando.

```
copy customer
from 's3://<your-bucket-name>/load/customer-fw.tbl'
credentials 'aws_iam_role=arn:aws:iam::<aws-account-id>:role/<role-name>'
fixedwidth 'c_custkey:10, c_name:25, c_address:25, c_city:10, c_nation:15,
 c_region :12, c_phone:15,c_mktsegment:10';
```

Debería obtener un mensaje de error, similar al siguiente.

```
An error occurred when executing the SQL command:
copy customer
from 's3://amzn-s3-demo-bucket/load/customer-fw.tbl'
credentials'...

ERROR: Load into table 'customer' failed.  Check 'stl_load_errors' system table for
 details. [SQL State=XX000]

Execution time: 2.95s

1 statement(s) failed.
```

MAXERROR

De forma predeterminada, la primera vez que COPY encuentra un error, el comando falla y devuelve un mensaje de error. Para ahorrar tiempo durante las pruebas, puede utilizar la opción MAXERROR para indicar a COPY que omita un número específico de errores antes de fallar. Dado que se esperan errores la primera vez que se prueba la carga de los datos de la tabla CUSTOMER, agregue `maxerror 10` al comando COPY.

Para realizar la prueba utilizando las opciones FIXEDWIDTH y MAXERROR, ejecute el siguiente comando.

```
copy customer
from 's3://<your-bucket-name>/load/customer-fw.tbl'
credentials 'aws_iam_role=arn:aws:iam::<aws-account-id>:role/<role-name>'
```

```
fixedwidth 'c_custkey:10, c_name:25, c_address:25, c_city:10, c_nation:15,
 c_region :12, c_phone:15,c_mktsegment:10'
maxerror 10;
```

Esta vez, en lugar de un mensaje de error, obtendrá un mensaje de advertencia, similar al siguiente.

```
Warnings:
Load into table 'customer' completed, 112497 record(s) loaded successfully.
Load into table 'customer' completed, 7 record(s) could not be loaded.  Check
 'stl_load_errors' system table for details.
```

La advertencia indica que COPY encontró siete errores. Para comprobar los errores, consulte la tabla STL_LOAD_ERRORS, como se muestra en el siguiente ejemplo.

```
select query, substring(filename,22,25) as filename,line_number as line,
substring(colname,0,12) as column, type, position as pos, substring(raw_line,0,30) as
 line_text,
substring(raw_field_value,0,15) as field_text,
substring(err_reason,0,45) as error_reason
from stl_load_errors
order by query desc, filename
limit 7;
```

Los resultados de la consulta a STL_LOAD_ERRORS deben ser similares a los siguientes.

```
 query  |         filename          | line |  column   |    type    | pos |
 line_text           | field_text |              error_reason
--------+---------------------------+------+-----------
+------------+-----+-------------------------------+------------
+--------------------------------------------
 334489 | customer-fw.tbl.log       |    2 | c_custkey | int4       |  -1 | customer-
fw.tbl               | customer-f | Invalid digit, Value 'c', Pos 0, Type: Integ
 334489 | customer-fw.tbl.log       |    6 | c_custkey | int4       |  -1 | Complete
                  | Complete   | Invalid digit, Value 'C', Pos 0, Type: Integ
 334489 | customer-fw.tbl.log       |    3 | c_custkey | int4       |  -1 | #Total rows
                  | #Total row | Invalid digit, Value '#', Pos 0, Type: Integ
 334489 | customer-fw.tbl.log       |    5 | c_custkey | int4       |  -1 | #Status
                  | #Status    | Invalid digit, Value '#', Pos 0, Type: Integ
 334489 | customer-fw.tbl.log       |    1 | c_custkey | int4       |  -1 | #Load file
                  | #Load file | Invalid digit, Value '#', Pos 0, Type: Integ
 334489 | customer-fw.tbl000        |    1 | c_address | varchar    |  34 | 1
 Customer#000000001  | .Mayag.ezR | String contains invalid or unsupported UTF8
```

```
 334489 | customer-fw.tbl000        |    1 | c_address | varchar    |  34 | 1
 Customer#000000001  | .Mayag.ezR | String contains invalid or unsupported UTF8
(7 rows)
```

Al examinar los resultados, podrá ver que hay dos mensajes en la columna `error_reasons`:

- ```
  Invalid digit, Value '#', Pos 0, Type: Integ
  ```

  El archivo `customer-fw.tbl.log` provoca estos errores. El problema es que este es un archivo de registro, no de datos, y no debe cargarse. Puede utilizar el archivo de manifiesto para evitar la carga del archivo incorrecto.

- ```
  String contains invalid or unsupported UTF8
  ```

  El tipo de datos VARCHAR es compatible con caracteres multibyte UTF-8 de hasta tres bytes. Si los datos de carga contienen caracteres no válidos o no compatibles, puede utilizar la opción ACCEPTINVCHARS para reemplazar cada carácter no válido por un carácter alternativo especificado.

Otro problema con la carga que es más difícil de detectar es cuando la carga produjo resultados inesperados. Para investigar este problema, ejecute el siguiente comando para consultar la tabla CUSTOMER.

```
select c_custkey, c_name, c_address
from customer
order by c_custkey
limit 10;
```

```
 c_custkey |          c_name           |         c_address
-----------+---------------------------+---------------------------
         2 | Customer#000000002        | XSTf4,NCwDVaWNe6tE
         2 | Customer#000000002        | XSTf4,NCwDVaWNe6tE
         3 | Customer#000000003        | MG9kdTD
         3 | Customer#000000003        | MG9kdTD
         4 | Customer#000000004        | XxVSJsL
         4 | Customer#000000004        | XxVSJsL
         5 | Customer#000000005        | KvpyuHCplrB84WgAi
         5 | Customer#000000005        | KvpyuHCplrB84WgAi
         6 | Customer#000000006        | sKZz0CsnMD7mp4Xd0YrBvx
```

```
         6 | Customer#000000006        | sKZz0CsnMD7mp4Xd0YrBvx
(10 rows)
```

Las filas deben ser únicas, pero están duplicadas.

Otra forma de comprobar los resultados inesperados es verificar el número de filas que se cargaron. En nuestro caso, deberían haberse cargado 100 000 filas, pero el mensaje de carga notificó que se cargaron 112 497 registros. Las filas adicionales se cargaron porque COPY cargo un archivo extraño, `customer-fw.tbl0000.bak`.

En este ejercicio, se utiliza un archivo de manifiesto para evitar la carga de los archivos incorrectos.

ACCEPTINVCHARS

De forma predeterminada, cuando COPY encuentra un carácter que no es compatible con los tipos de datos de la columna, omite la fila y devuelve un error. Para obtener más información acerca de los caracteres UTF-8 no válidos, consulte Errores de carga de caracteres multibyte.

Puede utilizar la opción MAXERRORS para ignorar los errores y continuar la carga, luego consulte STL_LOAD_ERRORS para ubicar los caracteres no válidos y después corrija los archivos de datos. No obstante, MAXERRORS es más adecuado para solucionar problemas de carga y en general no debe utilizarse en un entorno de producción.

La opción ACCEPTINVCHARS suele ser una mejor elección para administrar caracteres no válidos. ACCEPTINVCHARS le indica a COPY reemplazar cada carácter no válido por un carácter válido especificado y continuar con la operación de carga. Puede especificar cualquier carácter ASCII válido, excepto NULL, como el carácter de sustitución. El carácter de sustitución predeterminado es un signo de interrogación (?). COPY reemplaza los caracteres multibyte con una cadena de sustitución de igual longitud. Por ejemplo, un carácter de 4 bytes se reemplazaría con `'????'`.

COPY devuelve el número de filas que contenían caracteres UTF-8 no válidos. Además agrega una entrada a la tabla de sistema STL_REPLACEMENTS para cada fila afectada, hasta un máximo de 100 filas por sector del nodo. También se sustituyen los caracteres UTF-8 no válidos adicionales, pero esos eventos de sustitución no se registran.

ACCEPTINVCHARS es válido solo para las columnas VARCHAR.

En este paso, se agrega el ACCEPTINVCHARS con el carácter de reemplazo `'^'`.

MANIFEST

Cuando utiliza el comando COPY desde Amazon S3 mediante un prefijo de clave, existe el riesgo de cargar tablas no deseadas. Por ejemplo, la carpeta `'s3://amzn-s3-demo-bucket/load/` contiene ocho archivos de datos que comparten el prefijo de clave `customer-fw.tbl`, `customer-fw.tbl0000`, `customer-fw.tbl0001` y así sucesivamente. No obstante, la misma carpeta también contiene los archivos extraños `customer-fw.tbl.log` y `customer-fw.tbl-0001.bak`.

Para garantizar que carga todos los archivos correctos, y únicamente los archivos correctos, utilice un archivo de manifiesto. El manifiesto es un archivo de texto en formato JSON que muestra explícitamente la clave de objeto única para cada archivo de origen a cargar. Los objetos de archivos pueden encontrarse en diferentes carpetas o diferentes buckets, pero deben estar en la misma región. Para obtener más información, consulte MANIFEST.

A continuación se muestra el texto `customer-fw-manifest`.

```
{
  "entries": [
    {"url":"s3://<your-bucket-name>/load/customer-fw.tbl-000"},
    {"url":"s3://<your-bucket-name>/load/customer-fw.tbl-001"},
    {"url":"s3://<your-bucket-name>/load/customer-fw.tbl-002"},
    {"url":"s3://<your-bucket-name>/load/customer-fw.tbl-003"},
    {"url":"s3://<your-bucket-name>/load/customer-fw.tbl-004"},
    {"url":"s3://<your-bucket-name>/load/customer-fw.tbl-005"},
    {"url":"s3://<your-bucket-name>/load/customer-fw.tbl-006"},
    {"url":"s3://<your-bucket-name>/load/customer-fw.tbl-007"}
    ]
}
```

Para cargar los datos para la tabla CUSTOMER utilizando un archivo de manifiesto

1. Abra el archivo `customer-fw-manifest` en un editor de texto.
2. Sustituya *<your-bucket-name>* por el nombre del bucket.
3. Guarde el archivo.
4. Cargue el archivo a la carpeta de carga en el bucket.
5. Ejecute el siguiente comando COPY.

```
copy customer from 's3://<your-bucket-name>/load/customer-fw-manifest'
credentials 'aws_iam_role=arn:aws:iam::<aws-account-id>:role/<role-name>'
fixedwidth 'c_custkey:10, c_name:25, c_address:25, c_city:10, c_nation:15,
 c_region :12, c_phone:15,c_mktsegment:10'
maxerror 10
```

```
acceptinvchars as '^'
manifest;
```

### Carga de la tabla DWDATE con DATEFORMAT

En este paso, se utilizan las opciones DELIMITER y DATEFORMAT para cargar la tabla DWDATE .

Cuando se cargan las columnas DATE y TIMESTAMP, COPY espera el formato predeterminado, que es AAAA-MM-DD para fechas y AAAA-MM-DD HH:MM:SS para marcas temporales. Si los datos de carga no utilizan un formato predeterminado, puede utilizar DATEFORMAT y TIMEFORMAT para especificar el formato.

El siguiente fragmento muestra formatos de fecha en la tabla DWDATE. Observe que los formatos de fecha en la columna dos son inconsistentes.

```
19920104 1992-01-04          Sunday  January 1992 199201 Jan1992 1 4 4 1...
19920112 January 12, 1992 Monday  January 1992 199201 Jan1992 2 12 12 1...
19920120 January 20, 1992 Tuesday     January 1992 199201 Jan1992 3 20 20 1...
```

#### DATEFORMAT

Puede especificar solo un formato de fecha. Si los datos de carga contienen formatos inconsistentes, posiblemente en diferentes columnas, o si en el tiempo de carga no se conoce el formato, utilice DATEFORMAT con el argumento `'auto'`. Cuando se especifica `'auto'`, COPY reconoce cualquier formato válido de fecha u hora y lo convierte al formato predeterminado. La opción `'auto'` reconoce varios formatos que no son compatibles cuando se utiliza una cadena de DATEFORMAT y TIMEFORMAT. Para obtener más información, consulte Utilización del reconocimiento automático con DATEFORMAT y TIMEFORMAT.

Para cargar la tabla DWDATE, ejecute el siguiente comando COPY.

```
copy dwdate from 's3://<your-bucket-name>/load/dwdate-tab.tbl'
credentials 'aws_iam_role=arn:aws:iam::<aws-account-id>:role/<role-name>'
delimiter '\t'
dateformat 'auto';
```

### Carga de varios archivos de datos

Puede usar las opciones GZIP y COMPUPDATE para cargar una tabla.

Puede cargar una tabla desde un único archivo de datos o desde varios archivos. Esto le permite comparar los tiempos de carga de ambos métodos.

GZIP, LZOP y BZIP2

Puede comprimir los archivos utilizando los formatos de compresión gzip, lzop o bzip2. Cuando se carga desde archivos comprimidos, COPY descomprime los archivos durante el proceso de carga. La compresión de los archivos ahorra espacio de almacenamiento y reduce los tiempos de carga.

COMPUPDATE

Cuando COPY carga una tabla vacía sin codificaciones de compresión, analiza los datos de carga para determinar las codificaciones óptimas. Luego modifica la tabla para utilizar esas codificaciones antes de comenzar la carga. Este proceso de análisis toma tiempo, pero ocurre una sola vez por tabla, como máximo. Para ahorrar tiempo, puede omitir este paso desactivando COMPUPDATE. Para habilitar una evaluación precisa de los tiempos de COPY, se desactiva COMPUPDATE en este paso.

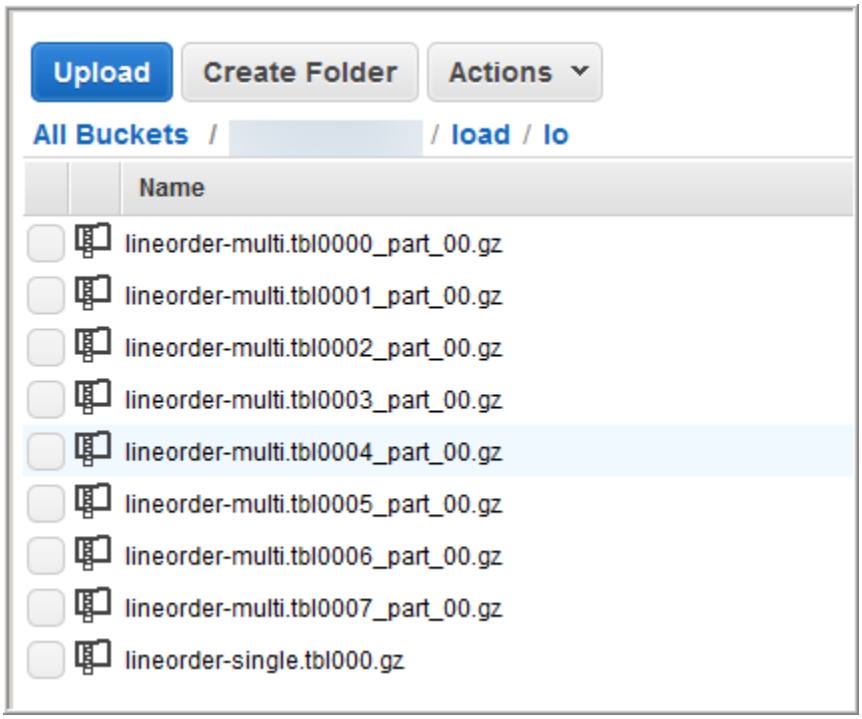
Múltiples archivos

El comando COPY puede cargar datos desde varios archivos en paralelo de manera muy eficiente, en lugar de desde un archivo individual. Puede dividir los datos en archivos de manera tal que la cantidad de archivos sea múltiplo de la cantidad de sectores en su clúster. Si lo hace, Amazon Redshift divide la carga de trabajo y distribuye los datos de manera uniforme entre los sectores. El número de sectores por nodo depende del tamaño de nodo del clúster. Para obtener más información acerca de la cantidad de sectores que tiene cada tamaño de nodo, consulte Acerca de clústeres y nodos en la Guía de administración de Amazon Redshift.

Por ejemplo, los nodos de computación del clúster que se utiliza en este tutorial tienen dos sectores cada uno, por lo que un clúster de cuatro nodos tendrá ocho sectores. En los pasos anteriores, los datos de carga estaban contenidos en ocho archivos, aunque los archivos sean muy pequeños. Puede comparar la diferencia de tiempo entre cargar desde un solo archivo grande y cargar desde varios archivos.

Incluso los archivos que contienen 15 millones de registros y ocupan alrededor de 1,2 GB son muy pequeños en la escala de Amazon Redshift, pero son suficientes para demostrar la ventaja de rendimiento que supone cargar desde varios archivos.

En la siguiente imagen se muestran los archivos de datos para LINEORDER.




Para evaluar el rendimiento de COPY con varios archivos

1. En una prueba de laboratorio, se ejecutó el siguiente comando para realizar la copia desde un único archivo. Este comando muestra un bucket ficticio.

```
copy lineorder from 's3://amzn-s3-demo-bucket/load/lo/lineorder-single.tbl'
credentials 'aws_iam_role=arn:aws:iam::<aws-account-id>:role/<role-name>'
gzip
compupdate off
region 'us-east-1';
```

2. Los resultados fueron los siguientes. Observe el tiempo de ejecución.

```
Warnings:
Load into table 'lineorder' completed, 14996734 record(s) loaded successfully.

0 row(s) affected.
copy executed successfully

Execution time: 51.56s
```

3. Luego, se ejecutó el siguiente comando para realizar la copia desde varios archivos.

```
copy lineorder from 's3://amzn-s3-demo-bucket/load/lo/lineorder-multi.tbl'
```

```
credentials 'aws_iam_role=arn:aws:iam::<aws-account-id>:role/<role-name>'
gzip
compupdate off
region 'us-east-1';
```

4. Los resultados fueron los siguientes. Observe el tiempo de ejecución.

```
Warnings:
Load into table 'lineorder' completed, 14996734 record(s) loaded successfully.

0 row(s) affected.
copy executed successfully

Execution time: 17.7s
```

5. Compare los tiempos de ejecución.

   En nuestro experimento, el tiempo para cargar 15 millones de registros se redujo de 51,56 segundos a 17,7 segundos, lo que supone una disminución del 65,7 %.

   Estos resultados se basan en la utilización de un clúster de cuatro nodos. Si el clúster tiene más nodos, el ahorro de tiempo se multiplica. Para los clústeres de Amazon Redshift típicos, que tienen de decenas a cientos de nodos, la diferencia es aún más grande. Si tiene un clúster de un nodo, hay una diferencia pequeña entre los tiempos de ejecución.

## Paso 6: Limpiar y analizar la base de datos

Siempre que agregue, elimine o modifique una cantidad importante de filas, debe ejecutar un comando VACUUM y luego un comando ANALYZE. Una limpieza recupera el espacio de las filas eliminadas y restaura el orden. El comando ANALYZE actualiza los metadatos estadísticos, lo que permite al optimizador de consultas generar planes de consultas más precisos. Para obtener más información, consulte Limpieza de tablas.

Si carga los datos en orden de clave de ordenación, la limpieza es rápida. En este tutorial, agregó un número importante de filas, pero lo hizo en tablas vacías. Siendo ese el caso, no hay necesidad de reordenar y no borró ninguna fila. COPY actualiza automáticamente las estadísticas después de cargar una tabla vacía, por lo que las estadísticas deben estar actualizadas. No obstante, por una cuestión de buena organización, se completa este tutorial con la limpieza y el análisis de la base de datos.

Para limpiar y analizar la base de datos, ejecute los siguientes comandos.

```
vacuum;
analyze;
```

## Paso 7: Eliminar los recursos

El clúster seguirá acumulando cargos mientras esté en ejecución. Una vez que haya completado este tutorial, debe restablecer el entorno al estado anterior si sigue los pasos que se indican en Paso 5: Revocar el acceso y eliminar el clúster de muestra en la Guía de introducción a Amazon Redshift.

Si desea mantener el clúster y recuperar el almacenamiento utilizado por las tablas de SSB, ejecute los siguientes comandos.

```
drop table part;
drop table supplier;
drop table customer;
drop table dwdate;
drop table lineorder;
```

### Next

Resumen

## Resumen

En este tutorial, ha cargado archivos de datos en Amazon S3 y ha utilizado los comandos COPY para cargar los datos de los archivos en tablas de Amazon Redshift.

Cargó datos utilizando los siguientes formatos:

- Delimitado por caracteres
- CSV
- De ancho fijo

Utilizó la tabla del sistema STL_LOAD_ERRORS para solucionar errores de carga y luego utilizó las opciones REGION, MANIFEST, MAXERROR, ACCEPTINVCHARS, DATEFORMAT y NULL AS para resolver los errores.

Aplicó las siguientes prácticas recomendadas para cargar datos:

- Uso del comando COPY para cargar datos
- Dividir los datos de carga
- Uso de un único comando COPY para cargar desde archivos múltiples
- Compresión de los archivos de datos
- Control de los archivos de datos antes y después de una carga

Para obtener más información acerca de las prácticas recomendadas de Amazon Redshift, consulte los siguientes enlaces:

- Prácticas recomendadas de Amazon Redshift para la carga de datos
- Prácticas recomendadas de Amazon Redshift para el diseño de tablas
- Prácticas recomendadas de Amazon Redshift para el diseño de consultas

# Descarga de datos en Amazon Redshift

Para descargar datos de las tablas de una base de datos en un conjunto de archivos de un bucket de Amazon S3, puede utilizar el comando UNLOAD con una instrucción SELECT. Puede descargar datos de texto en formato delimitado o formato de ancho fijo, independientemente del formato de datos que se utilizó para cargarlos. También puede especificar si desea crear archivos GZIP comprimidos.

Puede utilizar credenciales de seguridad temporales para limitar el acceso que tienen los usuarios al bucket de Amazon S3.

Temas


- Descarga de datos en Amazon S3
- Descarga de archivos de datos cifrados
- Descarga de datos en formato de ancho fijo o delimitado
- Volver a cargar datos descargados


## Descarga de datos en Amazon S3

Amazon Redshift divide los resultados de una instrucción de selección en un conjunto de archivos, uno o más archivos por sector de nodo, para simplificar la recarga paralela de los datos. De manera alternativa, puede especificar que UNLOAD debe escribir los resultados en serie en uno o más archivos al añadir la opción PARALLEL OFF. Puede especificar el parámetro MAXFILESIZE para limitar el tamaño de los archivos en Amazon S3. El comando UNLOAD cifra automáticamente los archivos de datos mediante el cifrado del lado del servidor de Amazon S3 (SSE-S3).

Puede utilizar cualquier instrucción de selección en el comando UNLOAD que sea compatible con Amazon Redshift, excepto una selección que utilice una cláusula LIMIT en la selección externa. Por ejemplo, puede utilizar una instrucción de selección que incluya columnas específicas o que use una cláusula WHERE para combinar varias tablas. Si la consulta contiene comillas (para delimitar valores literales, por ejemplo), debe aplicarle escape en el texto de la consulta (\'). Para obtener más información, consulte la referencia de comandos SELECT. Para obtener más información acerca del uso de una cláusula LIMIT, consulte Notas de uso para el comando UNLOAD.

Por ejemplo, el siguiente comando UNLOAD envía el contenido de la tabla VENUE al bucket `s3://amzn-s3-demo-bucket/tickit/unload/` de Amazon S3.

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/tickit/unload/venue_'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole';
```

Los nombres de los archivos creados por el ejemplo anterior incluyen el prefijo "venue_".

```
venue_0000_part_00
venue_0001_part_00
venue_0002_part_00
venue_0003_part_00
```

Por defecto, el comando UNLOAD escribe datos simultáneamente en varios archivos, según la cantidad de sectores que haya en el clúster. Para escribir datos en un único archivo, especifique PARALLEL OFF. El comando UNLOAD escribe los datos en serie, ordenados absolutamente en función de la cláusula ORDER BY, si se utiliza una. El tamaño máximo de un archivo de datos es de 6,2 GB. Si el tamaño de los datos es superior al máximo, UNLOAD crea archivos adicionales de hasta 6,2 GB cada uno.

En el siguiente ejemplo se escribe el contenido de VENUE en un único archivo. Solo se requiere un archivo debido a que el tamaño del archivo es menor a 6,2 GB.

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/tickit/unload/venue_'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
parallel off;
```

> **Note**
>
> El comando UNLOAD está diseñado para utilizar procesamiento en paralelo. Le recomendamos dejar PARALLEL habilitado para la mayoría de casos, especialmente si los archivos se utilizarán para cargar tablas mediante un comando COPY.

Suponiendo que el tamaño total de los datos de VENUE es 5 GB, en el siguiente ejemplo se escribe el contenido de VENUE en 50 archivos, cada uno de 100 MB.

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/tickit/unload/venue_'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
```

```
parallel off
maxfilesize 100 mb;
```

Si incluye un prefijo en la cadena de ruta de Amazon S3, el comando UNLOAD utilizará ese prefijo para los nombres de los archivos.

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/tickit/unload/venue_'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole';
```

Puede crear un archivo de manifiesto que enumere los archivos de descarga al especificar la opción MANIFEST en el comando UNLOAD. El manifiesto es un archivo de texto en formato JSON que enumera explícitamente la dirección URL de cada archivo escrito en Amazon S3.

En el siguiente ejemplo se incluye la opción de manifiesto.

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/tickit/venue_'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
manifest;
```

En el siguiente ejemplo se muestra un manifiesto para cuatro archivos de descarga.

```
{
  "entries": [
    {"url":"s3://amzn-s3-demo-bucket/tickit/venue_0000_part_00"},
    {"url":"s3://amzn-s3-demo-bucket/tickit/venue_0001_part_00"},
    {"url":"s3://amzn-s3-demo-bucket/tickit/venue_0002_part_00"},
    {"url":"s3://amzn-s3-demo-bucket/tickit/venue_0003_part_00"}
  ]
}
```

El archivo de manifiesto puede utilizarse para cargar los mismos archivos a través del comando COPY con la opción MANIFEST. Para obtener más información, consulte Uso de un manifiesto para especificar archivos de datos.

Después de completar una operación UNLOAD, confirme que los datos se descargaron de manera correcta al navegar hacia el bucket de Amazon S3 donde UNLOAD escribió los archivos. Verá uno o más archivos numerados por sector, comenzando con el número cero. Si especificó la opción MANIFEST, también verá un archivo que finaliza con "`manifest`". Por ejemplo:

```
amzn-s3-demo-bucket/tickit/venue_0000_part_00
amzn-s3-demo-bucket/tickit/venue_0001_part_00
amzn-s3-demo-bucket/tickit/venue_0002_part_00
amzn-s3-demo-bucket/tickit/venue_0003_part_00
amzn-s3-demo-bucket/tickit/venue_manifest
```

Puede obtener mediante programación una lista de los archivos que se escribieron en Amazon S3 llamando a una operación de lista de Amazon S3 una vez que se complete el comando UNLOAD. También puede consultar STL_UNLOAD_LOG.

La siguiente consulta devuelve el nombre de ruta para los archivos que se crearon a través de un comando UNLOAD. La función PG_LAST_QUERY_ID devuelve la consulta más reciente.

```
select query, substring(path,0,40) as path
from stl_unload_log
where query=2320
order by path;

query |             path
-------+---------------------------------------
  2320 | s3://amzn-s3-demo-bucket/venue0000_part_00
  2320 | s3://amzn-s3-demo-bucket/venue0001_part_00
  2320 | s3://amzn-s3-demo-bucket/venue0002_part_00
  2320 | s3://amzn-s3-demo-bucket/venue0003_part_00
(4 rows)
```

Si la cantidad de datos es muy grande, Amazon Redshift podría dividir los archivos en varias partes por sector. Por ejemplo:

```
venue_0000_part_00
venue_0000_part_01
venue_0000_part_02
venue_0001_part_00
venue_0001_part_01
venue_0001_part_02
...
```

El siguiente comando UNLOAD incluye una cadena entre comillas en la instrucción de selección, por lo que las comillas tienen aplicado un carácter de escape (=\'OH\' ').

```
unload ('select venuename, venuecity from venue where venuestate=\'OH\' ')
```

```
to 's3://amzn-s3-demo-bucket/tickit/venue/ '
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole';
```

Por defecto, UNLOAD producirá un error en lugar de sobrescribir los archivos existentes en el bucket de destino. Para sobrescribir los archivos existentes, incluido el archivo de manifiesto, especifique la opción ALLOWOVERWRITE.

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/venue_pipe_'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
manifest
allowoverwrite;
```

## Descarga de archivos de datos cifrados

El comando UNLOAD crea archivos automáticamente mediante el cifrado del servidor de Amazon S3 con claves de cifrado administradas por AWS (SSE-S3). También puede especificar cifrado del lado del servidor con una clave de AWS Key Management Service (SSE-KMS) o cifrado del lado del cliente con una clave administrada por el cliente. El comando UNLOAD no admite el cifrado de servidor de Amazon S3 utilizando una clave proporcionada por el cliente. Para obtener más información, consulte Protección de los datos con el cifrado del servidor.

Para descargar datos en Amazon S3 mediante el cifrado del lado del servidor con una clave de AWS KMS, utilice el parámetro KMS_KEY_ID para proporcionar el ID de la clave, tal como se muestra en el siguiente ejemplo.

```
unload ('select venuename, venuecity from venue')
to 's3://amzn-s3-demo-bucket/encrypted/venue_'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
KMS_KEY_ID '1234abcd-12ab-34cd-56ef-1234567890ab'
encrypted;
```

Si desea proporcionar su propia clave de cifrado, puede crear archivos de datos cifrados del lado del cliente en Amazon S3 mediante el comando UNLOAD con la opción ENCRYPTED. El comando UNLOAD utiliza el mismo proceso de cifrado de sobre que usa el cifrado del lado del cliente de Amazon S3. Puede utilizar el comando COPY con la opción ENCRYPTED para cargar los archivos cifrados.

El proceso ocurre de la siguiente manera:

1. Cree una clave de cifrado AES de 256 bits con codificación base64 que utilizará como clave de cifrado privada o clave simétrica principal.
2. Emita un comando UNLOAD que incluya la clave simétrica principal y la opción ENCRYPTED.
3. El comando UNLOAD genera una clave simétrica de un solo uso (denominada clave simétrica de sobre) y un vector de inicialización (IV), que se utiliza para cifrar los datos.
4. El comando UNLOAD cifra la clave simétrica de sobre a través de la clave simétrica principal.
5. El comando UNLOAD luego almacena los archivos de datos cifrados en Amazon S3 y almacena la clave de sobre cifrada y el IV como metadatos de los objetos en cada archivo. La clave de sobre cifrada se almacena como metadatos de los objetos `x-amz-meta-x-amz-key` y el IV se almacena como metadatos de los objetos `x-amz-meta-x-amz-iv`.

Para obtener más información acerca del proceso de cifrado de sobre, consulte Cifrado de datos del lado del cliente con el AWS SDK para Java y Amazon S3.

Para descargar archivos de datos cifrados, agregue el valor de la clave principal a la cadena de credenciales e incluya la opción ENCRYPTED. Si utiliza la opción MANIFEST, también se cifra el archivo de manifiesto.

```
unload ('select venuename, venuecity from venue')
to 's3://amzn-s3-demo-bucket/encrypted/venue_'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
master_symmetric_key '<root_key>'
manifest
encrypted;
```

Para descargar archivos de datos cifrados que están comprimidos en GZIP, incluya la opción GZIP junto con el valor de la clave principal y la opción ENCRYPTED.

```
unload ('select venuename, venuecity from venue')
to 's3://amzn-s3-demo-bucket/encrypted/venue_'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
master_symmetric_key '<root_key>'
encrypted gzip;
```

Para cargar archivos de datos cifrados, agregue el parámetro MASTER_SYMMETRIC_KEY con el mismo valor de clave principal e incluya la opción ENCRYPTED.

```
copy venue from 's3://amzn-s3-demo-bucket/encrypted/venue_'
```

```
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
master_symmetric_key '<root_key>'
encrypted;
```

## Descarga de datos en formato de ancho fijo o delimitado

Puede descargar datos en formato delimitado o formato de ancho fijo. El resultado predeterminado está delimitado por una barra vertical (con el carácter "|").

En el siguiente ejemplo se especifica una coma como delimitador:

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/tickit/venue/comma'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
delimiter ',';
```

Los archivos de salida resultantes tendrán este aspecto:

```
20,Air Canada Centre,Toronto,ON,0
60,Rexall Place,Edmonton,AB,0
100,U.S. Cellular Field,Chicago,IL,40615
200,Al Hirschfeld Theatre,New York City,NY,0
240,San Jose Repertory Theatre,San Jose,CA,0
300,Kennedy Center Opera House,Washington,DC,0
...
```

Para descargar el mismo conjunto de resultados a un archivo delimitado por tabulaciones, emita el siguiente comando:

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/tickit/venue/tab'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
delimiter as '\t';
```

O bien, puede usar una especificación FIXEDWIDTH. Esta especificación consta de un identificador para cada columna de la tabla y el ancho de la columna (cantidad de caracteres). El comando UNLOAD producirá un error en lugar de truncar los datos, por lo que debe especificar un ancho que tenga al menos la misma longitud que la entrada más larga de esa columna. Descargar datos de ancho fijo funciona de manera similar a descargar datos delimitados, excepto que la salida resultante no contiene caracteres que delimitan. Por ejemplo:

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/tickit/venue/fw'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
fixedwidth '0:3,1:100,2:30,3:2,4:6';
```

La salida de ancho fijo tendrá este aspecto:

```
20 Air Canada Centre         Toronto      ON0
60 Rexall Place              Edmonton     AB0
100U.S. Cellular Field       Chicago      IL40615
200Al Hirschfeld Theatre     New York CityNY0
240San Jose Repertory TheatreSan Jose     CA0
300Kennedy Center Opera HouseWashington   DC0
```

Para obtener más información acerca de las especificaciones FIXEDWIDTH, consulte el comando UNLOAD.

## Volver a cargar datos descargados

Para volver a cargar los resultados de una operación de descarga, puede usar el comando COPY.

En el siguiente ejemplo se muestra un caso simple en el que la tabla VENUE se descarga con un archivo de manifiesto, se trunca y se vuelve a cargar.

```
unload  ('select * from venue order by venueid')
to 's3://amzn-s3-demo-bucket/tickit/venue/reload_'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
manifest
delimiter '|';

truncate venue;

copy venue
from 's3://amzn-s3-demo-bucket/tickit/venue/reload_manifest'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
manifest
delimiter '|';
```

Después de volver a cargarla, la tabla VENUE tendrá este aspecto:

```
select * from venue order by venueid limit 5;

 venueid |         venuename          |  venuecity  | venuestate | venueseats
---------+----------------------------+-------------+------------+-----------
       1 | Toyota Park                | Bridgeview  | IL         |          0
       2 | Columbus Crew Stadium      | Columbus    | OH         |          0
       3 | RFK Stadium                | Washington  | DC         |          0
       4 | CommunityAmerica Ballpark | Kansas City | KS         |          0
       5 | Gillette Stadium           | Foxborough  | MA         |      68756
(5 rows)
```

# Funciones definidas por el usuario en Amazon Redshift

Puede crear una nueva función escalar definida por el usuario (UDF) personalizada mediante una cláusula SQL SELECT o un programa Python. La función nueva se almacena en la base de datos y está disponible para cualquier usuario que tenga los privilegios necesarios para ejecutarla. Una UDF escalar personalizada se ejecuta de la misma manera que se ejecutan las funciones existentes de Amazon Redshift.

Para las UDF de Python, además de usar la funcionalidad estándar en Python, puede importar sus propios módulos personalizados desde Python. Para obtener más información, consulte Compatibilidad del lenguaje Python con las UDF. Tenga en cuenta que Python 3 no está disponible para las UDF de Python. Para obtener compatibilidad con Python 3 para las UDF de Amazon Redshift, utilice UDF de Lambda escalares en su lugar.

También puede crear UDF de AWS Lambda que utilizan funciones personalizadas definidas en Lambda como parte de sus consultas SQL. Las UDF de Lambda permiten la escritura de UDF complejas y la integración a componentes de terceros. También pueden ayudar a superar algunas de las limitaciones de las UDF actuales de Python y SQL. Por ejemplo, pueden ayudar a acceder a los recursos de red y almacenamiento, además de a escribir instrucciones SQL más completas. Puede crear UDF de Lambda en cualquiera de los lenguajes de programación admitidos por Lambda, como Java, Go, PowerShell, Node.js, C#, Python y Ruby. También puede utilizar un tiempo de ejecución personalizado.

De manera predeterminada, todos los usuarios pueden ejecutar UDF. Para obtener más información acerca de los privilegios, consulte Seguridad y permisos de UDF.

Temas


- Seguridad y permisos de UDF
- Prevención de conflictos al dar nombre a las UDF
- UDF de SQL escalares
- UDF de Python escalares
- UDF de Lambda escalares
- Ejemplos de casos de uso para funciones definidas por el usuario (UDF)

# Seguridad y permisos de UDF

Para crear una UDF, debe tener permiso para usar el lenguaje de SQL o plpythonu (Python). De manera predeterminada, a USAGE ON LANGUAGE SQL se le concede el permiso PUBLIC, pero debe conceder explícitamente el permiso de USAGE ON LANGUAGE PLPYTHONU a usuarios o grupos específicos.

Para revocar los permisos de uso de SQL, primero revoque el uso de PUBLIC. A continuación, conceda los permisos de uso de SQL a solo los usuarios o grupos que tienen permiso para crear UDF de SQL. El siguiente ejemplo revoca el uso de SQL de PUBLIC. A continuación, concede permisos de uso al grupo de usuarios `udf_devs`.

```
revoke usage on language sql from PUBLIC;
grant usage on language sql to group udf_devs;
```

Para ejecutar una UDF, debe tener los permisos correspondientes para cada función. De manera predeterminada, el permiso para ejecutar nuevas UDF está establecido en PUBLIC. Para restringir el uso, revoque el permiso PUBLIC para la función. A continuación, conceda el privilegio a usuarios o grupos específicos.

El siguiente ejemplo revoca la ejecución de la función `f_py_greater` de PUBLIC. A continuación, concede permisos de uso al grupo de usuarios `udf_devs`.

```
revoke execute on function f_py_greater(a float, b float) from PUBLIC;
grant execute on function f_py_greater(a float, b float) to group udf_devs;
```

De manera predeterminada, los superusuarios tienen todos los privilegios.

Para obtener más información, consulte GRANT y REVOKE.

# Prevención de conflictos al dar nombre a las UDF

Puede evitar resultados inesperados y conflictos potenciales si considera las convenciones para dar nombre a la User-Defined Scalar Function (UDF, Función escalar definida por el usuario) antes de la implementación. Como los nombres de las funciones pueden sobrecargarse, pueden confundirse con nombres de funciones de Amazon Redshift existentes y futuras. En este tema, se analiza la sobrecarga y se presenta una estrategia para evitar cualquier conflicto.

## Sobrecarga de los nombres de función

Una función se identifica por su nombre y su firma, que es la cantidad de argumentos de entrada y los tipos de datos de los argumentos. Dos funciones del mismo esquema pueden tener el mismo nombre si tienen diferentes firmas. En otras palabras, los nombres de funciones pueden sobrecargarse.

Cuando se ejecuta una consulta, el motor de consultas determina qué función se debe llamar en función de la cantidad de argumentos proporcionados y los tipos de datos de esos argumentos. Puede utilizar la sobrecarga para simular funciones con una cantidad variable de argumentos, hasta el límite permitido por el comando CREATE FUNCTION.

## Cómo evitar conflictos con las funciones integradas de Amazon Redshift

Recomendamos que los nombres que asigne a las UDF incluyan el prefijo `f_`. Amazon Redshift reserva el prefijo `f_` exclusivamente para las UDF y, al utilizar el prefijo `f_` en los nombres de las UDF, se asegura de que estos no entren en conflicto con los nombres de las funciones SQL futuras o existentes integradas de Amazon Redshift. Por ejemplo, si asigna el nombre `f_sum` a una UDF nueva, evitará conflictos con la función SUM de Amazon Redshift. De igual modo, si asigna el nombre `f_fibonacci` a una función nueva, evitará conflictos en el caso de que Amazon Redshift agregue una función llamada FIBONACCI en una versión futura.

Puede crear una UDF con el mismo nombre y la misma firma que una función SQL integrada de Amazon Redshift existente sin que se sobrecargue el nombre de la función, cada vez que la UDF y la función integrada existan en esquemas diferentes. Como las funciones integradas existen en el esquema de catálogo del sistema, pg_catalog, puede crear una UDF con el mismo nombre en otro esquema; por ejemplo, uno público o definido por el usuario. En algunos casos, es posible que llame a una función que no esté calificada explícitamente con un nombre de esquema. Si es así, Amazon Redshift busca primero el esquema pg_catalog de manera predeterminada. Por lo tanto, una función integrada se ejecuta antes que una UDF nueva con el mismo nombre.

Puede cambiar este comportamiento si configura la ruta de búsqueda de manera que coloque pg_catalog al final. Si lo hace, las UDF tendrán prioridad sobre las funciones integradas, pero esta práctica puede generar resultados inesperados. La adopción de una estrategia de nombres única, como el uso del prefijo reservado `f_`, es una práctica más confiable. Para obtener más información, consulte SET y search_path.

# UDF de SQL escalares

Una UDF de SQL escalar incorpora una cláusula SQL SELECT que se ejecuta cuando se llama a la función y se devuelve un solo valor. El comando CREATE FUNCTION define los siguientes parámetros:

- Argumentos de entrada (opcionales). Cada argumento debe tener un tipo de datos.
- Un tipo de datos de retorno.
- Una cláusula SQL SELECT. En la cláusula SELECT, haga referencia a los argumentos de entrada usando $1, $2, etc., en función del orden de los argumentos en la definición de la función.

Los tipos de datos de entrada y de retorno pueden ser cualquier tipo de dato de Amazon Redshift estándar.

No incluya una cláusula FROM en la cláusula SELECT. En lugar de ello, incluya la cláusula FROM en la instrucción SQL que llama a la UDF de SQL.

La cláusula SELECT no puede incluir ninguno de los tipos de cláusulas siguientes:

- FROM
- INTO
- WHERE
- GROUP BY
- ORDER BY
- LIMIT

## Ejemplo de una función SQL escalar

En el siguiente ejemplo, se crea una función que compara dos números y devuelve el valor más grande. Para obtener más información, consulte CREATE FUNCTION.

```
create function f_sql_greater (float, float)
  returns float
stable
as $$
  select case when $1 > $2 then $1
```

```
    else $2
  end
$$ language sql;
```

La siguiente consulta llama a la nueva función f_sql_greater para consultar la tabla SALES y devolver COMMISSION o el 20 por ciento de PRICEPAID, lo que sea mayor.

```
select f_sql_greater(commission, pricepaid*0.20) from sales;
```

## UDF de Python escalares

Una UDF de Python escalar incorpora un programa Python que se ejecuta cuando se llama a la función y se obtiene un solo valor. El comando CREATE FUNCTION define los siguientes parámetros:

- Argumentos de entrada (opcionales). Cada argumento debe tener un nombre y un tipo de datos.
- Un tipo de datos de retorno.
- Un programa en Python ejecutable.

Los tipos de datos de entrada y devolución para las UDF de Python pueden ser cualquiera de los siguientes tipos:

- SMALLINT
- INTEGER
- BIGINT
- DECIMAL
- REAL
- DOUBLE PRECISION
- BOOLEAN
- CHAR
- VARCHAR
- DATE
- TIMESTAMP
- ANYELEMENT

Los alias para estos tipos también son válidos. Para obtener una lista completa de tipos de datos y los alias, consulte Tipos de datos.

Cuando las UDF de Python utilizan el tipo de datos ANYELEMENT, Amazon Redshift lo convierte automáticamente en un tipo de datos estándar basado en los argumentos proporcionados en tiempo de ejecución. Para obtener más información, consulte Tipo de datos ANYELEMENT.

Los siguientes pasos se producen durante el tiempo de ejecución cuando una consulta de Amazon Redshift llama a una UDF escalar:

1. La función convierte los argumentos de entrada a tipos de datos de Python.

   Para la asignación de los tipos de datos de Amazon Redshift y los tipos de datos de Python, consulte Tipos de datos de UDF de Python.
2. La función ejecuta el programa en Python y pasa los argumentos de entrada convertidos.
3. El código en Python devuelve un valor único. El tipo de dato del valor de retorno debe coincidir con el tipo de dato RETURNS especificado por la definición de la función.
4. La función convierte el valor de retorno en Python al tipo de datos de Amazon Redshift especificado y, luego, devuelve ese valor a la consulta.

> **Note**
>
> Python 3 no está disponible para las UDF de Python. Para obtener compatibilidad con Python 3 para las UDF de Amazon Redshift, utilice UDF de Lambda escalares en su lugar.

## Ejemplo de una UDF escalar de Python

En el siguiente ejemplo, se crea una función que compara dos números y devuelve el valor más grande. Tenga en cuenta que la sangría del código entre los signos de dólar dobles ($$) es un requisito de Python. Para obtener más información, consulte CREATE FUNCTION.

```
create function f_py_greater (a float, b float)
  returns float
stable
as $$
  if a > b:
    return a
  return b
```

```
$$ language plpythonu;
```

La siguiente consulta solicita la nueva función `f_greater` para consultar la tabla SALES y devolver COMMISSION o 20 por ciento de PRICEPAID, cualquiera que represente el mayor valor.

```
select f_py_greater (commission, pricepaid*0.20) from sales;
```

## Tipos de datos de UDF de Python

Las UDF de Python pueden utilizar cualquier tipo de datos estándar de Amazon Redshift para los argumentos de entrada y el valor de retorno de la función. Además de los tipos de datos estándar, las UDF admiten el tipo de datos ANYELEMENT, que Amazon Redshift convierte de forma automática a un tipo de datos estándar en función de los argumentos proporcionados durante el tiempo de ejecución. Las UDF escalares pueden devolver un tipo de datos de ANYELEMENT. Para obtener más información, consulte Tipo de datos ANYELEMENT.

Durante la ejecución, Amazon Redshift convierte los argumentos de tipos de datos de Amazon Redshift a tipos de datos de Python para su procesamiento. Luego, convierte el valor de retorno del tipo de datos de Python al tipo de datos de Amazon Redshift correspondiente. Para obtener más información acerca de los tipos de datos de Amazon Redshift, consulte Tipos de datos.

En la siguiente tabla, se relacionan los tipos de datos de Amazon Redshift y los de Python.

| Tipos de datos de Amazon Redshift | Tipo de datos de Python |
|---|---|
| smallint | int |
| integer | |
| bigint | |
| short | |
| long | |
| decimal o numeric | decimal |
| double | float |
| real | |

| Tipos de datos de Amazon Redshift | Tipo de datos de Python |
|---|---|
| booleano | bool |
| char<br>varchar | cadena |
| marca de tiempo | datetime |

### Tipo de datos ANYELEMENT

ANYELEMENT es un tipo de datos polimórfico. Esto significa que, si una función se declara con ANYELEMENT para el tipo de datos de un argumento, la función puede aceptar cualquier tipo de datos estándar de Amazon Redshift como entrada para ese argumento cuando se llama a la función. El argumento ANYELEMENT se establece según el tipo de datos que, de hecho, pasan a él cuando se solicita la función.

Si una función usa distintos tipos de datos ANYELEMENT, todos deben resolverse según el mismo tipo de datos real cuando se solicita la función. Todos los tipos de datos de los argumentos ANYELEMENT se establecen según el tipo de datos real del primer argumento que pasa a un ANYELEMENT. Por ejemplo, una función declarada como `f_equal(anyelement, anyelement)` tomará dos valores de entrada cualquiera, siempre que sean del mismo tipo de datos.

Si el valor de retorno de una función se declara como ANYELEMENT, al menos un argumento de entrada debe ser en ANYELEMENT. El tipo de datos real del valor de retorno es el mismo que el tipo de datos real proporcionado para el argumento de entrada ANYELEMENT.

## Compatibilidad del lenguaje Python con las UDF

Puede crear una User-Defined Scalar Function (UDF, Función escalar definida por el usuario) personalizada basada en el lenguaje de programación Python. La biblioteca estándar de Python 2.7 se encuentra disponible para utilizarse en UDF, a excepción de los siguientes módulos:

- ScrolledText
- Tix
- Tkinter
- tk

- turtle
- smtpd

Además de la biblioteca estándar de Python, los siguientes módulos forman parte de la implementación de Amazon Redshift:

- numpy 1.8.2
- pandas 0.14.1
- python-dateutil 2.2
- pytz 2014.7
- scipy 0.12.1
- six 1.3.0
- wsgiref 0.1.2

También puede importar sus propios módulos personalizados desde Python y hacer que estén disponibles para su uso en UDF mediante la ejecución de un comando CREATE LIBRARY. Para obtener más información, consulte Ejemplo: Importación de módulos personalizados desde la biblioteca de Python.

> **Important**
>
> Amazon Redshift bloquea el acceso a la red y el acceso de escritura al sistema de archivos a través de las UDF.

> **Note**
>
> Python 3 no está disponible para las UDF de Python. Para obtener compatibilidad con Python 3 para las UDF de Amazon Redshift, utilice UDF de Lambda escalares en su lugar.

## Ejemplo: Importación de módulos personalizados desde la biblioteca de Python

Define las funciones escalares con sintaxis del lenguaje Python. Puede utilizar los módulos de la biblioteca estándar de Python y los módulos preinstalados de Amazon Redshift. También puede

crear sus propios módulos personalizados de la biblioteca de Python e importar las bibliotecas a sus clústeres, o utilizar las bibliotecas existentes de Python o terceros.

No puede crear una biblioteca que contenga un módulo con el mismo nombre que un módulo de la biblioteca estándar de Python o un módulo de Python preinstalado de Amazon Redshift. Si una biblioteca existente instalada por el usuario usa el mismo paquete de Python que la biblioteca que creó, debe eliminar la biblioteca existente antes de instalar la nueva biblioteca.

Para instalar bibliotecas personalizadas debe ser un super usuario o tener el privilegio `USAGE ON LANGUAGE plpythonu`; no obstante, cualquier usuario con privilegios suficientes como para crear funciones puede usar las bibliotecas instaladas. Puede consultar el catálogo del sistema PG_LIBRARY para obtener información acerca de las bibliotecas instaladas en su clúster.

Importación de un módulo personalizado de Python al clúster

En la presente sección, se ofrece un ejemplo de importación de un módulo personalizado de Python a su clúster. Para llevar a cabo los pasos de esta sección, debe tener un bucket de Amazon S3 donde cargar el paquete de la biblioteca. Luego instalará el paquete en su clúster. Para obtener más información acerca de la creación de buckets, consulte Creación de un bucket en la Guía del usuario de Amazon Simple Storage Service.

En este ejemplo, supongamos que crea una UDF para trabajar con posiciones y distancias en sus datos. Conéctese a su clúster de Amazon Redshift desde una herramienta cliente SQL y ejecute los siguientes comandos para crear las funciones.

```
CREATE FUNCTION f_distance (x1 float, y1 float, x2 float, y2 float) RETURNS float
 IMMUTABLE as $$
    def distance(x1, y1, x2, y2):
        import math
        return math.sqrt((y2 - y1) ** 2 + (x2 - x1) ** 2)

    return distance(x1, y1, x2, y2)
$$ LANGUAGE plpythonu;

CREATE FUNCTION f_within_range (x1 float, y1 float, x2 float, y2 float) RETURNS bool
 IMMUTABLE as $$
    def distance(x1, y1, x2, y2):
        import math
        return math.sqrt((y2 - y1) ** 2 + (x2 - x1) ** 2)

    return distance(x1, y1, x2, y2) < 20
```

```
$$ LANGUAGE plpythonu;
```

Tenga en cuenta que algunas líneas de código se duplican en las funciones previas. Esa duplicación es necesaria porque una UDF no puede hacer referencia al contenido de otra UDF y ambas funciones requieren la misma funcionalidad. No obstante, en lugar de duplicar el código en varias funciones, puede crear una biblioteca personalizada y configurar las funciones para que la usen.

Para ello, en primer lugar cree el paquete de biblioteca siguiendo los pasos a continuación:

1. Cree una carpeta con el nombre geometry. Esa carpeta es el paquete de nivel superior de la biblioteca.
2. En la carpeta geometry, cree un archivo llamado `__init__.py`. Tenga en cuenta que el nombre del archivo tiene dos caracteres de guión bajo doble. Este archivo le indica a Python que el paquete puede inicializarse.
3. En la carpeta geometry, cree también una carpeta llamada trig. Esa carpeta es el subpaquete de la biblioteca.
4. En la carpeta trig, cree otro archivo llamado `__init__.py` y un archivo llamado `line.py`. En esta carpeta, `__init__.py` le indica a Python que el subpaquete puede inicializarse y que `line.py` es el archivo que tiene el código de biblioteca.

   La estructura de carpetas y archivos debe ser igual a la siguiente:

   ```
   geometry/
      __init__.py
      trig/
         __init__.py
         line.py
   ```

   Para obtener más información acerca de la estructura del paquete, visite Modules en el tutorial de Python que se encuentra en el sitio web de Python.
5. El siguiente código tiene una clase y las funciones de miembro de la biblioteca. Cópielo y péguelo en `line.py`.

   ```
   class LineSegment:
     def __init__(self, x1, y1, x2, y2):
       self.x1 = x1
       self.y1 = y1
       self.x2 = x2
       self.y2 = y2
   ```

```
  def angle(self):
    import math
    return math.atan2(self.y2 - self.y1, self.x2 - self.x1)
  def distance(self):
    import math
    return math.sqrt((self.y2 - self.y1) ** 2 + (self.x2 - self.x1) ** 2)
```

Una vez que haya creado el paquete, haga lo siguiente para prepararlo y cargarlo en Amazon S3.

1. Comprima el contenido de la carpeta geometry en un archivo .zip con el nombre geometry.zip. No incluya la carpeta geometry en sí; incluya solamente el contenido de la carpeta, tal y como se muestra a continuación:

```
geometry.zip
   __init__.py
   trig/
      __init__.py
      line.py
```

2. Cargue geometry.zip en el bucket de Amazon S3.

> **Important**
>
> Si el bucket de Amazon S3 no se encuentra en la misma región que el clúster de Amazon Redshift, debe utilizar la opción REGION para especificar la región en la que se encuentran los datos. Para obtener más información, consulte CREATE LIBRARY.

3. Desde su herramienta cliente SQL, ejecute el siguiente comando para instalar la biblioteca. Reemplace *<bucket_name>* con el nombre del bucket y reemplace *<access key id>* y *<secret key>* con una clave de acceso y una clave de acceso secreta de las credenciales de usuario de AWS Identity and Access Management (IAM).

```
CREATE LIBRARY geometry LANGUAGE plpythonu FROM 's3://<bucket_name>/geometry.zip'
 CREDENTIALS 'aws_access_key_id=<access key id>;aws_secret_access_key=<secret key>';
```

Después de instalar la biblioteca en su clúster, debe configurar sus funciones para poder usarla. Para ello, ejecute los siguientes comandos.

```
CREATE OR REPLACE FUNCTION f_distance (x1 float, y1 float, x2 float, y2 float) RETURNS
 float IMMUTABLE as $$
    from trig.line import LineSegment

    return LineSegment(x1, y1, x2, y2).distance()
$$ LANGUAGE plpythonu;

CREATE OR REPLACE FUNCTION f_within_range (x1 float, y1 float, x2 float, y2 float)
 RETURNS bool IMMUTABLE as $$
    from trig.line import LineSegment

    return LineSegment(x1, y1, x2, y2).distance() < 20
$$ LANGUAGE plpythonu;
```

En los comandos anteriores, `import trig/line` elimina el código duplicado de las funciones originales presentadas en esta sección. Puede reusar la funcionalidad brindada por esta biblioteca en distintas UDF. Tenga en cuenta que para importar el módulo, solamente necesita especificar la ruta del subpaquete y el nombre del módulo (`trig/line`).

## Restricciones de UDF de Python

Entre las restricciones mencionadas en este tema, puede utilizar las UDF en cualquier lugar en que use las funciones escalares integradas de Amazon Redshift. Para obtener más información, consulte Referencia de funciones SQL.

Las UDF de Python de Amazon Redshift presentan las siguientes restricciones:

- Las UDF de Python no pueden obtener acceso a la red ni leer o escribir en el sistema de archivos.
- El tamaño total de las bibliotecas de Python instaladas por el usuario no puede superar los 100 MB.
- Amazon Redshift solo puede ejecutar una UDF de Python a la vez para clústeres aprovisionados que utilicen la administración de cargas de trabajo (WLM) automática y para grupos de trabajo sin servidor. Si intenta ejecutar varias UDF de forma simultánea, Amazon Redshift pondrá en cola las UDF de Python restantes para que se ejecuten en las colas de administración de cargas de trabajo. Las UDF de SQL no tienen límite de simultaneidad cuando se utiliza la WLM automática.
- Cuando se utiliza la WLM manual para clústeres aprovisionados, el número de UDF de Python que pueden ejecutarse simultáneamente por clúster está limitado a una cuarta parte del nivel de simultaneidad total del clúster. Por ejemplo, un clúster aprovisionado con una simultaneidad de 15 puede ejecutar un máximo de tres UDF de Python simultáneas.

- Al utilizar UDF de Python, Amazon Redshift no admite los tipos de datos SUPER y HLLSKETCH.

## Registro de errores y advertencias en las UDF de Python

Puede usar el módulo de registro de Python para crear mensajes de error y advertencia definidos por el usuario en sus User-Defined Scalar Function (UDF, Función escalar definida por el usuario). Después de la ejecución de una consulta, puede consultar la vista de sistema SVL_UDF_LOG para recuperar mensajes registrados.

> **Note**
>
> El registro de UDF consume recursos del clúster y podría afectar el rendimiento del sistema. Le recomendamos que implemente el registro solo con fines de desarrollo y resolución de problemas.

Durante la ejecución de consultas, el gestor de registros escribe mensajes a la vista de sistema SVL_UDF_LOG, junto con la función, nombre, nodo y sector correspondientes. El gestor de registros escribe una fila en SVL_UDF_LOG por mensaje, por sector. Los mensajes se truncan en 4 096 bytes. El registro de UDF se limita a 500 filas por sector. Cuando el registro se llena, el gestor descarta los mensajes más viejos y agrega un mensaje de advertencia a SVL_UDF_LOG.

> **Note**
>
> El gestor de registros de las UDF de Amazon Redshift introduce los escapes de las líneas nuevas (\n), las barras verticales (|) y las barras inversas (\) con una barra inversa (\).

De manera predeterminada, el nivel registro de la UDF se establece en WARNING. Los mensajes con un nivel registro WARNING, ERROR y CRITICAL se registran. Los mensajes de menor gravedad, INFO, DEBUG y NOTSET, se ignoran. Para establecer el nivel registro de la UDF, use el método de registro en Python. Por ejemplo, lo siguiente establece el nivel registro en INFO.

```
logger.setLevel(logging.INFO)
```

Para obtener más información acerca del uso del módulo de registro en Python, consulte Logging facility for Python en la documentación de Python.

En el siguiente ejemplo, se crea una función denominada f_pyerror que importa el módulo de registro en Python, crea una instancia de registro y registra un error.

```
CREATE OR REPLACE FUNCTION f_pyerror()
RETURNS INTEGER
VOLATILE AS
$$
import logging

logger = logging.getLogger()
logger.setLevel(logging.INFO)
logger.info('Your info message here')
return 0
$$ language plpythonu;
```

En el siguiente ejemplo, se consulta el SVL_UDF_LOG para ver el mensaje registrado en el ejemplo anterior.

```
select funcname, node, slice, trim(message) as message
from svl_udf_log;

  funcname  | query | node | slice |   message
------------+-------+------+-------+-----------------
  f_pyerror | 12345 |     1|     1 | Your info message here
```

# UDF de Lambda escalares

Amazon Redshift puede utilizar funciones personalizadas definidas en AWS Lambda como parte de las consultas SQL. Puede escribir UDF de Lambda escalares en cualquier lenguaje de programación admitido por Lambda, como Java, Go, PowerShell, Node.js, C#, Python y Ruby. También puede utilizar un tiempo de ejecución personalizado.

El comando CREATE EXTERNAL FUNCTION crea los siguientes parámetros:

- (Opcional) Una lista de argumentos con tipo de datos.
- Un tipo de datos de retorno.
- Un nombre de función de la función externa a la que llama Amazon Redshift.
- Un rol de IAM que el clúster de Amazon Redshift está autorizado a asumir y llamar a Lambda.
- Un nombre de función de Lambda que invoca la UDF de Lambda.

Para obtener más información acerca de CREATE EXTERNAL FUNCTION, consulte CREATE EXTERNAL FUNCTION.

Los tipos de datos de entrada y devolución para las UDF de Lambda pueden ser cualquiera de los siguientes tipos:

- SMALLINT
- INTEGER
- BIGINT
- DECIMAL
- REAL
- DOUBLE PRECISION
- CHAR
- VARCHAR
- BOOLEAN
- DATE
- TIMESTAMP

Los alias para estos tipos también son válidos. Para obtener una lista completa de tipos de datos y los alias, consulte Tipos de datos.

Amazon Redshift se asegura de que la función externa pueda enviar y recibir argumentos y resultados en lotes.

Las UDF de Lambda se definen y administran en Lambda. Además, puede controlar los privilegios de acceso para invocar estas UDF en Amazon Redshift. Puede invocar varias funciones de Lambda en la misma consulta o invocar la misma función varias veces.

Utilice las UDF de Lambda en cualquier cláusula de las instrucciones SQL donde se admitan funciones escalares. También puede utilizar las UDF de Lambda en cualquier instrucción SQL, como SELECT, UPDATE, INSERT o DELETE.

> **Note**
>
> El uso de las UDF de Lambda puede generar cargos adicionales por el servicio Lambda. Si esto sucede o no depende de distintos factores, como el número de solicitudes de Lambda (invocaciones de UDF) y la duración total de la ejecución del programa Lambda. No obstante,

no se generan cargos adicionales por el uso de las UDF de Lambda en Amazon Redshift. Para obtener información acerca de los precios de AWS Lambda, consulte Precios de AWS Lambda.

El número de solicitudes de Lambda varía según la cláusula de instrucción SQL específica en la que se utiliza la UDF de Lambda. Por ejemplo, supongamos que la función se utiliza en una cláusula WHERE como la siguiente.

```
SELECT a, b FROM t1 WHERE lambda_multiply(a, b) = 64; SELECT a, b
FROM t1 WHERE a*b = lambda_multiply(2, 32)
```

En este caso, Amazon Redshift llama a la primera instrucción SELECT para cada una y llama a la segunda instrucción SELECT solo una vez.

No obstante, es posible que el uso de una UDF en la parte de proyección de la consulta solo invoque la función de Lambda una vez por cada fila calificada o agrupada en el conjunto de resultados.

## Seguridad y permisos de UDF

Para crear una UDF de Lambda, asegúrese de tener los permisos de uso en LANGUAGE EXFUNC. Debe conceder de forma explícita USAGE ON LANGUAGE EXFUNC o revocar USAGE ON LANGUAGE EXFUNC a usuarios, grupos o público específicos.

En el siguiente ejemplo, se establece el uso de EXFUNC como PUBLIC.

```
grant usage on language exfunc to PUBLIC;
```

En el siguiente ejemplo, se revoca el permiso de uso PUBLIC de exfunc y se concede el permiso de uso al grupo de usuarios lambda_udf_devs.

```
revoke usage on language exfunc from PUBLIC;
grant usage on language exfunc to group lambda_udf_devs;
```

Para ejecutar una UDF de Lambda, asegúrese de tener permiso para cada función a la que se llama. De manera predeterminada, se establece como PUBLIC el permiso para ejecutar las UDF de Lambda nuevas. Para restringir el uso, revoque el permiso PUBLIC para la función. Luego, conceda el privilegio a usuarios o grupos específicos.

En el siguiente ejemplo, se revoca la ejecución de la función exfunc_sum como PUBLIC. Luego, se concede el uso al grupo de usuarios lambda_udf_devs.

```
revoke execute on function exfunc_sum(int, int) from PUBLIC;
grant execute on function exfunc_sum(int, int) to group lambda_udf_devs;
```

De manera predeterminada, los superusuarios tienen todos los privilegios.

Para obtener más información sobre cómo conceder y revocar privilegios, consulte GRANT y REVOKE.

## Configuración del parámetro de autorización para las UDF de Lambda

El comando CREATE EXTERNAL FUNCTION requiere autorización para invocar funciones de Lambda en AWS Lambda. Para comenzar la autorización, especifique un rol de AWS Identity and Access Management (IAM) cuando ejecute el comando CREATE EXTERNAL FUNCTION. Para obtener más información acerca de los roles de IAM, consulte Roles de IAM en la Guía del usuario de IAM.

Si ya existe un rol de IAM con permisos para invocar funciones de Lambda adjuntas a su clúster, puede sustituir el nombre de recurso de Amazon (ARN) del rol en el parámetro IAM_ROLE con el comando. En las siguientes secciones, se describen los pasos para utilizar un rol de IAM en el comando CREATE EXTERNAL FUNCTION.

### Creación de un rol de IAM para Lambda

El rol de IAM requiere permiso para invocar funciones de Lambda. Cuando cree el rol de IAM, proporcione el permiso a través de una de las siguientes maneras:

- Adjunte la política `AWSLambdaRole` en la página Attach permissions policy (Adjuntar política de permisos) cuando cree un rol de IAM. La política `AWSLambdaRole` concede permisos para invocar funciones de Lambda, que es el requisito mínimo. Para obtener más información y otras políticas, consulte Políticas de IAM basadas en identidad para AWS Lambda en la Guía para desarrolladores de AWS Lambda.
- Cree su propia política personalizada para adjuntarla a su rol de IAM con el permiso `lambda:InvokeFunction` de todos los recursos o de una función de Lambda en particular con el ARN de esa función. Para obtener más información sobre cómo crear una política, consulte Creación de políticas de IAM en la Guía del usuario de IAM.

  La siguiente política de ejemplo permite invocar a Lambda en una función de Lambda en particular.

  ```
  {
  ```

```
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "Invoke",
            "Effect": "Allow",
            "Action": [
                "lambda:InvokeFunction"
            ],
            "Resource": "arn:aws:lambda:us-west-2:123456789012:function:my-function"
        }
    ]
}
```

Para obtener más información sobre los recursos para funciones de Lambda, consulte Recursos y condiciones para acciones de Lambda en la Referencia de la API de IAM.

Después de crear una política personalizada con los permisos necesarios, puede adjuntar la política al rol de IAM en la página Attach permissions policy (Adjuntar política de permisos) cuando cree un rol de IAM.

Para conocer los pasos que se deben realizar para crear un rol de IAM, consulte Autorización de Amazon Redshift para acceder a otros servicios de AWS en su nombre en la Guía de administración de Amazon Redshift.

Si no desea crear un rol de IAM nuevo, puede agregar los permisos mencionados anteriormente a su rol de IAM existente.

### Asociación de un rol de IAM al clúster

Adjunte el rol de IAM a su clúster. Puede agregar un rol a un clúster o ver los roles asociados a un clúster mediante la consola de administración, la CLI o la API de Amazon Redshift. Para obtener más información, consulte Asociación de un rol de IAM a un clúster en la Guía de administración de Amazon Redshift.

### Inclusión del rol de IAM en el comando

Incluya el ARN del rol de IAM en el comando CREATE EXTERNAL FUNCTION. Al crear un rol de IAM, IAM devuelve el Amazon Resource Name (ARN, Nombre de recurso de Amazon) del rol. Para especificar un rol de IAM, proporcione el ARN del rol con el parámetro `IAM_ROLE`. A continuación, se muestra la sintaxis del parámetro `IAM_ROLE`.

```
IAM_ROLE 'arn:aws:iam::aws-account-id:role/role-name'
```

Para invocar funciones de Lambda que se encuentran en otras cuentas dentro de la misma región, consulte Encadenamiento de roles de IAM en Amazon Redshift.

## Uso de la interfaz JSON entre Amazon Redshift y AWS Lambda

Amazon Redshift utiliza una interfaz común para todas las funciones de Lambda con las que Amazon Redshift se comunica.

La siguiente tabla muestra la lista de campos de entrada con las funciones designadas de Lambda que puede esperar con la carga de JSON.

| Nombre del campo | Descripción | Rango de valor |
|---|---|---|
| request_id | Un identificador único universal (UUID) que identifica de manera inequívoca cada solicitud de invocación | Un UUID válido |
| clúster | El nombre de recurso de Amazon (ARN) completo del clúster | Un ARN de clúster válido |
| usuario | El nombre del usuario que realiza la llamada | Un nombre de usuario válido |
| database | El nombre de la base de datos en la que se ejecuta la consulta | Un nombre de base de datos válido |
| external_function | El nombre válido y completo de la función externa que realiza la llamada | Un nombre válido y completo de función |

| Nombre del campo | Descripción | Rango de valor |
|---|---|---|
| query_id | El ID de la consulta que realiza la llamada | Un ID de consulta válido |
| num_records | El número de argumentos de la carga | Un valor de 1 - 2^64 |
| arguments | La carga de datos en el formato especificado | Los datos en formato de matriz deben estar en una matriz JSON. Cada elemento es un registro que es una matriz si el número de argumentos es mayor que 1. Mediante el uso de una matriz, Amazon Redshift conserva el orden de los registros de la carga. |

El orden de la matriz JSON determina el orden del procesamiento por lotes. La función de Lambda debe procesar los argumentos de manera iterativa y producir el número exacto de registros. A continuación, se muestra un ejemplo de carga.

```
{
  "request_id" : "23FF1F97-F28A-44AA-AB67-266ED976BF40",
  "cluster" : "arn:aws:redshift:xxxx",
  "user" : "adminuser",
  "database" : "db1",
  "external_function": "public.foo",
  "query_id" : 5678234,
  "num_records" : 4,
  "arguments" : [
     [ 1, 2 ],
     [ 3, null],
     null,
     [ 4, 6]
   ]
 }
```

La salida de la función de Lambda contiene los siguientes campos.

| Nombre del campo | Descripción | Rango de valor |
|---|---|---|
| success | La indicación del éxito o del error de la función | Un valor de `"true"` o `"false"` |
| error_msg | El mensaje de error si el valor de éxito se establece como `"false"` (si la función presenta error); de lo contrario, este campo se ignora. | Un mensaje válido |
| num_records | El número de registros de la carga | Un valor de 1 - 2^64 |
| results | Los resultados de la llamada en el formato especificado | N/A |

El siguiente es un ejemplo de la salida de la función de Lambda.

```
{
  "success": true,   // true indicates the call succeeded
  "error_msg" : "my function isn't working",  // shall only exist when success != true
  "num_records": 4,      // number of records in this payload
  "results" : [
     1,
     4,
     null,
     7
   ]
}
```

Cuando llama a las funciones de Lambda desde consultas SQL, Amazon Redshift garantiza la seguridad de la conexión con las siguientes consideraciones:

- Permisos GRANT y REVOKE. Para obtener más información acerca de los permisos y la seguridad de UDF, consulte Seguridad y permisos de UDF.
- Amazon Redshift solo envía el conjunto mínimo de datos a la función de Lambda designada.
- Amazon Redshift solo llama a la función de Lambda designada con el rol de IAM designado.

# Ejemplos de casos de uso para funciones definidas por el usuario (UDF)

Puede utilizar funciones definidas por el usuario para resolver problemas empresariales integrando Amazon Redshift con otros componentes. A continuación se incluyen algunos ejemplos de cómo otros usuarios han utilizado UDF para sus casos de uso:

- Acceso a componentes externos mediante UDF de Lambda de Amazon Redshift: describe cómo funcionan las UDF de Lambda de Amazon Redshift y explica cómo crear una UDF de Lambda.
- Traducción y análisis de texto mediante funciones SQL con Amazon Redshift, Amazon Translate y Amazon Comprehend: proporciona UDF de Lambda de Amazon Redshift prediseñadas que se pueden instalar con unos pocos clics para traducir, redactar y analizar campos de texto.
- Acceso a Amazon Location Service desde Amazon Redshift: describe cómo utilizar las UDF de Lambda de Amazon Redshift para integrarse con Amazon Location Service.
- Tokenización de datos con Amazon Redshift y Protegrity: describe cómo integrar las UDF de Lambda de Amazon Redshift con el producto Protegrity sin servidor.
- UDF de Amazon Redshift: una colección de UDF de Amazon Redshift SQL, Lambda y Python.

# Creación de procedimientos almacenados en Amazon Redshift

En este tema se describe cómo crear y utilizar procedimientos almacenados en Amazon Redshift. Un procedimiento almacenado es un conjunto de sentencias SQL que pueden utilizar varios programas.

Puede definir un procedimiento almacenado de Amazon Redshift con el lenguaje de procedimientos PL/pgSQL de PostgreSQL para realizar un conjunto de operaciones lógicas y consultas de SQL. El procedimiento se almacena en la base de datos y está disponible para cualquier usuario con los suficientes privilegios de base de datos.

A diferencia de las funciones definidas por el usuario (UDF), un procedimiento almacenado puede incorporar lenguaje de definición de datos (DDL) y lenguaje de manipulación de datos (DML) además de consultas SELECT. Un procedimiento almacenado no tiene que devolver un valor. Puede utilizar lenguaje de procedimientos, incluidos bucles y expresiones condicionales, para controlar el flujo lógico.

Para obtener más detalles de los comandos de SQL para crear y gestionar procedimientos almacenados, consulte los siguientes temas:

- CREATE PROCEDURE
- ALTER PROCEDURE
- DROP PROCEDURE
- SHOW PROCEDURE
- CALL
- GRANT
- REVOKE
- ALTER DEFAULT PRIVILEGES

Temas

- Información general de procedimientos almacenados en Amazon Redshift
- Referencia del lenguaje PL/pgSQL

# Información general de procedimientos almacenados en Amazon Redshift

En este tema se describen los detalles sobre la finalidad y el uso de los procedimientos almacenados.

Los procedimientos almacenados se utilizan normalmente para encapsular lógica destinada a la transformación de datos, la validación de datos y la lógica específica para empresas. Mediante la combinación de varios pasos de SQL en un procedimiento almacenado, puede reducir las idas y vueltas entre sus aplicaciones y la base de datos.

Para lograr un control de acceso preciso, puede crear procedimientos almacenados que ejecuten funciones sin otorgar a un usuario acceso a las tablas subyacentes. Por ejemplo, solo el propietario o un superusuario pueden truncar una tabla. Un usuario necesita privilegios de escritura para insertar datos en una tabla. En lugar de conceder privilegios al usuario para las tablas subyacentes, puede crear un procedimiento almacenado que realice la tarea. A continuación, puede dar privilegios al usuario para que ejecute el procedimiento almacenado.

Un procedimiento almacenado con el atributo de seguridad DEFINER se ejecuta con los privilegios del propietario del procedimiento almacenado. De forma predeterminada, un procedimiento almacenado tiene seguridad INVOKER, lo que significa que el procedimiento usa los privilegios del usuario que llama el procedimiento.

Para crear un procedimiento almacenado, use el comando CREATE PROCEDURE. Para ejecutar un procedimiento, use el comando CALL. Se proporcionan ejemplos más adelante en esa sección.

> **Note**
>
> Algunos clientes pueden mostrar el siguiente error al crear un procedimiento almacenado de Amazon Redshift.
>
> ```
> ERROR: 42601: [Amazon](500310) unterminated dollar-quoted string at or near "$$
> ```
>
> Este error ocurre debido a la incapacidad del cliente para analizar correctamente la instrucción CREATE PROCEDURE con puntos y coma delimitando las instrucciones y con un símbolo del dólar ($) citado. Esto causa que solo una parte de la instrucción se envíe al servidor de Amazon Redshift. A menudo, puede sortear este error usando las opciones `Run as batch` o `Execute selected` del cliente.

Por ejemplo, si usa un cliente Aginity, use la opción `Run entire script as batch`. Si usa SQL Workbench/J, recomendamos la versión 124. Si usa la versión 125 de SQL Workbench/J, considere especificar un delimitador alternativo como solución. CREATE PROCEDURE contiene instrucciones SQL delimitadas por un punto y coma (;). Al definir un delimitador alternativo, como una barra diagonal (/) y situarlo al final de una instrucción CREATE PROCEDURE, se envía la instrucción entera al servidor de Amazon Redshift para procesarla. A continuación se muestra un ejemplo.

```
CREATE OR REPLACE PROCEDURE test()
AS $$
BEGIN
  SELECT 1 a;
END;
$$
LANGUAGE plpgsql
;
/
```

También puede usar un cliente con compatibilidad para analizar instrucciones CREATE PROCEDURE, como el editor de consultas en la consola de Amazon Redshift o TablePlus.

Temas

- Nomenclatura de los procedimientos almacenados
- Seguridad y privilegios para procedimientos almacenados
- Devolución de un conjunto de resultados desde un procedimiento almacenado
- Administración de transacciones
- Errores de captura
- Registro de procedimientos almacenados
- Limitaciones de los procedimientos almacenados

El siguiente ejemplo muestra un procedimiento sin argumentos de salida. De forma predeterminada, los argumentos son argumentos de entrada (IN).

```
CREATE OR REPLACE PROCEDURE test_sp1(f1 int, f2 varchar)
AS $$
BEGIN
```

```
  RAISE INFO 'f1 = %, f2 = %', f1, f2;
END;
$$ LANGUAGE plpgsql;

call test_sp1(5, 'abc');
INFO: f1 = 5, f2 = abc
CALL
```

> **Note**
>
> Al escribir procedimientos almacenados, sugerimos una práctica recomendada para proteger los valores confidenciales:
> No codifique ninguna información confidencial en la lógica de procedimientos almacenados. Por ejemplo, no asigne una contraseña de usuario en una instrucción CREATE USER del cuerpo de un procedimiento almacenado. Esto supone un riesgo de seguridad, porque los valores codificados de esta forma se pueden registrar como metadatos de esquema en las tablas del catálogo. En cambio, pase valores confidenciales, como las contraseñas, como argumentos al procedimiento almacenado, mediante parámetros.
> Para obtener más información sobre los procedimientos almacenados, consulte CREATE PROCEDURE y Creación de procedimientos almacenados en Amazon Redshift. Para obtener más información sobre las tablas de catálogo, consulte Tablas de catálogos de sistema.

El siguiente ejemplo muestra un procedimiento con argumentos de salida. Los argumentos son entrada (IN), entrada y salida (INOUT) y salida (OUT).

```
CREATE OR REPLACE PROCEDURE test_sp2(f1 IN int, f2 INOUT varchar(256), out_var OUT
 varchar(256))
AS $$
DECLARE
  loop_var int;
BEGIN
  IF f1 is null OR f2 is null THEN
    RAISE EXCEPTION 'input cannot be null';
  END IF;
  DROP TABLE if exists my_etl;
  CREATE TEMP TABLE my_etl(a int, b varchar);
    FOR loop_var IN 1..f1 LOOP
        insert into my_etl values (loop_var, f2);
```

```
        f2 := f2 || '+' || f2;
    END LOOP;
  SELECT INTO out_var count(*) from my_etl;
END;
$$ LANGUAGE plpgsql;


call test_sp2(2,'2019');

         f2          | column2
---------------------+---------
 2019+2019+2019+2019 | 2
(1 row)
```

## Nomenclatura de los procedimientos almacenados

En este tema se describen los detalles sobre los nombres de los procedimientos almacenados.

Si define un procedimiento con el mismo nombre y diferentes tipos de datos de argumentos, o con una firma diferente, creará un nuevo procedimiento. Como resultado, el nombre del procedimiento se sobrecarga. Para obtener más información, consulte Sobrecarga de nombres de procedimientos. Amazon Redshift no habilita la sobrecarga de procedimientos en función de argumentos de salida. No puede tener dos procedimientos con el mismo nombre y tipos de datos de argumentos de entrada y diferentes tipos de argumentos de salida.

El propietario o un superusuario puede sustituir el cuerpo de un procedimiento almacenado por otro nuevo con la misma firma. Para cambiar la firma o los tipos devueltos de un procedimiento almacenado, anule el procedimiento almacenado y vuelva a crearlo. Para obtener más información, consulte DROP PROCEDURE y CREATE PROCEDURE.

Puede evitar resultados inesperados y conflictos potenciales si analiza las convenciones de nomenclatura de los procedimientos almacenados antes de la implementación. Como se puede sobrecargar los nombres de los procedimientos, estos pueden entrar en conflicto con nombres de procedimientos de Amazon Redshift presentes y futuros.

### Sobrecarga de nombres de procedimientos

Un procedimiento se identifica por su nombre y su firma, que es la cantidad de argumentos de entrada y los tipos de datos de los argumentos. Dos procedimientos del mismo esquema pueden tener el mismo nombre si tienen diferentes firmas. En otras palabras, puede sobrecargar los nombres de los procedimientos.

Al ejecutar un procedimiento, el motor de consultas determina qué procedimiento debe llamar a partir de la cantidad de argumentos brindados y de los tipos de datos de esos argumentos. Puede usar la sobrecarga para simular procedimientos con una cantidad variable de argumentos, hasta el límite permitido por el comando CREATE PROCEDURE. Para obtener más información, consulte CREATE PROCEDURE.

### Prevención de conflictos de nomenclatura

Recomendamos que nombre todos los procedimientos usando el prefijo `sp_`. Amazon Redshift reserva el prefijo `sp_` exclusivamente para procedimientos almacenados. Si utiliza el prefijo `sp_` en los nombres de los procedimientos, se asegura de que el nombre no entre en conflicto con ningún nombre de procedimiento de Amazon Redshift existente o futuro.

## Seguridad y privilegios para procedimientos almacenados

En este tema se describen las credenciales de base de datos necesarias para crear y ejecutar procedimientos almacenados.

De forma predeterminada, todos los usuarios tienen privilegios para crear un procedimiento. Para crear un procedimiento, debe tener el privilegio USAGE en el lenguaje PL/pgSQL, que se concede a PUBLIC de forma predeterminada. Solo los superusuarios y propietarios tienen el privilegio para llamar a un procedimiento de forma predeterminada. Los superusuarios pueden ejecutar REVOKE USAGE en PL/pgSQL de un usuario si quieren impedir que el usuario cree procedimientos almacenados.

Para llamar a un procedimiento, se le debe haber concedido el privilegio EXECUTE para el procedimiento. De forma predeterminada, al propietario del procedimiento y a los superusuarios se les concede el privilegio EXECUTE para nuevos procedimientos. Para obtener más información, consulte GRANT.

El usuario que crea el procedimiento es el propietario de forma predeterminada. El propietario tiene privilegios CREATE, DROP y EXECUTE sobre el procedimiento de forma predeterminada. Los superusuarios tienen todos los privilegios.

El atributo SECURITY controla los privilegios de un procedimiento para acceder a objetos de la base de datos. Al crear un procedimiento almacenado, puede establecer el atributo SECURITY en DEFINER o INVOKER. Este atributo determina qué privilegios se utilizan al ejecutar las instrucciones en el cuerpo del procedimiento almacenado. Si especifica SECURITY INVOKER, el procedimiento usa los privilegios del usuario que invoca el procedimiento. Si especifica SECURITY DEFINER,

el procedimiento usa los privilegios del propietario del procedimiento. INVOKER es el valor predeterminado.

Como un procedimiento SECURITY DEFINER se ejecuta con los privilegios del usuario que es su propietario, debe asegurarse de que el procedimiento no se pueda usar indebidamente. Para asegurarse que los procedimientos SECURITY DEFINER no puedan usarse indebidamente, haga lo siguiente:

- Otorgue EXECUTE sobre procedimientos SECURITY DEFINER a usuarios concretos y no a PUBLIC.
- Cualifique todos los objetos de base de datos a los que el procedimiento debe acceder con los nombres de esquema. Por ejemplo, utilice `myschema.mytable` en lugar de `mytable`.
- Si no puede cualificar un nombre de objeto por su esquema, establezca `search_path` al crear el procedimiento mediante la opción SET. Establezca `search_path` para excluir los esquemas en los que pueden escribir los usuarios que no son de confianza. Esta estrategia impide que los intermediarios de este procedimiento creen objetos (como tablas o vistas) que enmascaren objetos destinados a ser usados por el procedimiento. Para obtener más información acerca de la opción SET, consulte CREATE PROCEDURE.

Los siguientes ejemplos establecen `search_path` en `admin` para asegurar que la tabla `user_creds` sea accesible desde el esquema `admin` y no desde el esquema público ni ningún otro esquema en la `search_path` del intermediario.

```
CREATE OR REPLACE PROCEDURE sp_get_credentials(userid int, o_creds OUT varchar)
AS $$
BEGIN
  SELECT creds INTO o_creds
  FROM user_creds
  WHERE user_id = $1;
END;
$$ LANGUAGE plpgsql
SECURITY DEFINER
-- Set a secure search_path
SET search_path = admin;
```

# Devolución de un conjunto de resultados desde un procedimiento almacenado

En este tema se describe cómo devuelven datos los procedimientos almacenados.

Puede devolver conjuntos de resultados mediante un cursor o una tabla temporal.

## Devolución de un cursor

Para devolver un cursor, cree un procedimiento con un argumento INOUT definido con un tipos de datos `refcursor`. Cuando llame al procedimiento, asigne un nombre al cursor. Después, puede obtener los resultados del cursor por nombre.

El siguiente ejemplo crea un procedimiento llamado `get_result_set` con un argumento INOUT llamado `rs_out` mediante el tipo de datos `refcursor`. El procedimiento abre el cursos mediante una instrucción SELECT.

```
CREATE OR REPLACE PROCEDURE get_result_set (param IN integer, rs_out INOUT refcursor)
AS $$
BEGIN
  OPEN rs_out FOR SELECT * FROM fact_tbl where id >= param;
END;
$$ LANGUAGE plpgsql;
```

El siguiente comando CALL abre el cursor con el nombre `mycursor`. Use cursores solo con transacciones.

```
BEGIN;
CALL get_result_set(1, 'mycursor');
```

Después de abrir el cursor, puede realizar recuperaciones desde el cursor, como se muestra en el siguiente ejemplo.

```
FETCH ALL FROM mycursor;

    id | secondary_id | name
-------+--------------+---------
     1 |            1 | Joe
     1 |            2 | Ed
     2 |            1 | Mary
```

```
     1 |            3 | Mike
(4 rows)
```

Al final, la transacción se confirma o revierte.

```
COMMIT;
```

Un cursor devuelto por un procedimiento almacenado está sujeto a las mismas limitaciones y consideraciones de rendimiento que se describen en DECLARE CURSOR. Para obtener más información, consulte Restricciones del cursor.

El siguiente ejemplo muestra la llamada del procedimiento almacenado `get_result_set` mediante un tipo de datos `refcursor` de JDBC. El nombre `'mycursor'` literal (el nombre del cursor) se pasa a `prepareStatement`. A continuación, los resultados se recuperan de `ResultSet`.

```
static void refcursor_example(Connection conn) throws SQLException {
    conn.setAutoCommit(false);
    PreparedStatement proc = conn.prepareStatement("CALL get_result_set(1,
 'mycursor')");
    proc.execute();
    ResultSet rs = statement.executeQuery("fetch all from mycursor");
    while (rs.next()) {
      int n = rs.getInt(1);
      System.out.println("n " + n);
    }
```

## Uso de una tabla temporal

Para devolver resultados, puede devolver un controlador a una tabla temporal que contiene filas de resultados. El cliente puede proporcionar un nombre como un parámetro para el procedimiento almacenado. Dentro del procedimiento almacenado, SQL dinámico puede usarse para operar en la tabla temporal. A continuación se muestra un ejemplo.

```
CREATE PROCEDURE get_result_set(param IN integer, tmp_name INOUT varchar(256)) as $$
DECLARE
  row record;
BEGIN
  EXECUTE 'drop table if exists ' || tmp_name;
  EXECUTE 'create temp table ' || tmp_name || ' as select * from fact_tbl where id <= '
 || param;
```

```
END;
$$ LANGUAGE plpgsql;

CALL get_result_set(2, 'myresult');
 tmp_name
-----------
 myresult
(1 row)

SELECT * from myresult;
 id | secondary_id | name
----+--------------+------
  1 |            1 | Joe
  2 |            1 | Mary
  1 |            2 | Ed
  1 |            3 | Mike
(4 rows)
```

## Administración de transacciones

Puede crear un procedimiento almacenado con un comportamiento de administración de transacciones predeterminado o un comportamiento no atómico.

### Administración de transacciones de procedimientos almacenados en modo predeterminado

El comportamiento de confirmación automática de modo de transacción predeterminado hace que cada comando de SQL ejecutado por separado se confirme individualmente. Una llamada a un procedimiento almacenado se trata como un comando de SQL sencillo. Las instrucciones de SQL dentro de un procedimiento se comportan como si estuvieran en una bloque de transacciones que, de forma implícita, empieza cuando se inicia la llamada y acaba cuando finaliza la llamada. Una llamada anidada en otro procedimiento se trata como cualquier otra instrucción de SQL y opera dentro del contexto de la misma transacción como el intermediario. Para obtener más información acerca del comportamiento de confirmación automática, consulte Aislamiento serializable.

No obstante, supongamos que llama a un procedimiento almacenado desde dentro de un bloque de transacciones especificado del usuario (definido por BEGIN…COMMIT). En este caso, todas las instrucciones del procedimiento almacenado se ejecutan en el contexto de la transacción especificada por el usuario. El procedimiento no se confirma de forma implícita en la salida. El intermediario controla la confirmación o reversión del procedimiento.

Si hay un error mientras se ejecuta un procedimiento almacenado, se revierten todos los cambios realizados en la transacción actual.

Puede usar las siguientes instrucciones de control de transacción en un proceso almacenado:

- COMMIT: confirma todo el trabajo realizado en la transacción actual y comienza una nueva transacción de forma implícita. Para obtener más información, consulte COMMIT.
- ROLLBACK: invierte el trabajo realizado en la transacción actual y comienza una nueva transacción de forma implícita. Para obtener más información, consulte ROLLBACK.

TRUNCATE es otra instrucción que puede surgir en un proceso almacenado e influye en la administración de las transacciones. En Amazon Redshift, TRUNCATE emite una confirmación de forma implícita. Este comportamiento es el mismo en el contexto de procedimientos almacenados. Cuando se emite una instrucción TRUNCATE desde dentro de un procedimiento almacenado, se confirma la transacción actual y empieza una nueva. Para obtener más información, consulte TRUNCATE.

Todas las instrucciones que siguen a una instrucción COMMIT, ROLLBACK o TRUNCATE funcionan en el contexto de una nueva transacción. Lo hacen hasta que se halla una instrucción COMMIT, ROLLBACK o TRUNCATE o hasta que el proceso almacenado sale.

Cuando usa COMMIT, ROLLBACK o TRUNCATE desde dentro de un procedimiento almacenado, se aplican las siguientes limitaciones:

- Si se llama al procedimiento almacenado desde un bloque de transacciones, no puede emitir una instrucción COMMIT, ROLLBACK o TRUNCATE. Esta restricción aplica dentro del propio cuerpo del procedimiento almacenado y dentro de cualquier llamada de un proceso anidado.
- Si se crea el procedimiento almacenado con opciones `SET config`, no puede emitir una instrucción COMMIT, ROLLBACK o TRUNCATE. Esta restricción aplica dentro del propio cuerpo del procedimiento almacenado y dentro de cualquier llamada de un proceso anidado.
- Cualquier cursor abierto (explícita o implícitamente) se cierra automáticamente cuando se procesa una instrucción COMMIT, ROLLBACK o TRUNCATE. Para conocer las limitaciones de cursores explícitas e implícitas, consulte Limitaciones de los procedimientos almacenados.

Adicionalmente, no puede ejecutar COMMIT o ROLLBACK usando SQL dinámico. Sin embargo, puede ejecutar TRUNCATE usando SQL dinámico. Para obtener más información, consulte SQL dinámico.

Cuando trabaje con procesos almacenados, considere que las instrucciones BEGIN y END en PL/pgSQL son solo para agrupar. No inician ni finalizan una transacción. Para obtener más información, consulte Bloque.

El siguiente ejemplo muestra el comportamiento de transacción cuando se llama a un procedimiento almacenado desde dentro de un bloque de transacciones explícitas. Las dos instrucciones de inserción emitidas desde fuera del procedimiento almacenado y la que se emite desde dentro forman parte de la misma transacción (3382). La transacción se confirma cuando el usuario emite una confirmación explícita.

```
CREATE OR REPLACE PROCEDURE sp_insert_table_a(a int) LANGUAGE plpgsql
AS $$
BEGIN
  INSERT INTO test_table_a values (a);
END;
$$;

Begin;
  insert into test_table_a values (1);
  Call sp_insert_table_a(2);
  insert into test_table_a values (3);
Commit;

select userid, xid, pid, type, trim(text) as stmt_text
from svl_statementtext where pid = pg_backend_pid() order by xid , starttime ,
 sequence;

 userid | xid  | pid |  type   |               stmt_text
--------+------+-----+---------+----------------------------------------
    103 | 3382 | 599 | UTILITY | Begin;
    103 | 3382 | 599 | QUERY   | insert into test_table_a values (1);
    103 | 3382 | 599 | UTILITY | Call sp_insert_table_a(2);
    103 | 3382 | 599 | QUERY   | INSERT INTO test_table_a values ( $1 )
    103 | 3382 | 599 | QUERY   | insert into test_table_a values (3);
    103 | 3382 | 599 | UTILITY | COMMIT
```

Por contraste, hay un ejemplo cuando las mismas instrucciones se emiten desde el exterior del bloque de una transacción específica y la sesión tiene la autoconfirmación establecida en ON. En este caso, cada instrucción se ejecuta en su propia transacción.

```
insert into test_table_a values (1);
Call sp_insert_table_a(2);
```

```
insert into test_table_a values (3);

select userid, xid, pid, type, trim(text) as stmt_text
from svl_statementtext where pid = pg_backend_pid() order by xid , starttime ,
 sequence;

 userid | xid  | pid |  type   |
             stmt_text
--------+------+-----+---------
+-----------------------------------------------------------------------------------------------
    103 | 3388 | 599 | QUERY   | insert into test_table_a values (1);
    103 | 3388 | 599 | UTILITY | COMMIT
    103 | 3389 | 599 | UTILITY | Call sp_insert_table_a(2);
    103 | 3389 | 599 | QUERY   | INSERT INTO test_table_a values ( $1 )
    103 | 3389 | 599 | UTILITY | COMMIT
    103 | 3390 | 599 | QUERY   | insert into test_table_a values (3);
    103 | 3390 | 599 | UTILITY | COMMIT
```

El siguiente ejemplo emite una instrucción TRUNCATE después de realizar una inserción en `test_table_a`. La instrucción TRUNCATE emite una confirmación implícita que confirma la transacción actual (3335) y empieza una nueva (3336). La nueva transacción se confirma cuando se produce la salida del procedimiento.

```
CREATE OR REPLACE PROCEDURE sp_truncate_proc(a int, b int) LANGUAGE plpgsql
AS $$
BEGIN
  INSERT INTO test_table_a values (a);
  TRUNCATE test_table_b;
  INSERT INTO test_table_b values (b);
END;
$$;

Call sp_truncate_proc(1,2);

select userid, xid, pid, type, trim(text) as stmt_text
from svl_statementtext where pid = pg_backend_pid() order by xid , starttime ,
 sequence;

 userid | xid  |  pid  |  type   |
                                        stmt_text
--------+------+-------+---------
+-----------------------------------------------------------------------------------------------
    103 | 3335 | 23636 | UTILITY | Call sp_truncate_proc(1,2);
```

```
    103 | 3335 | 23636 | QUERY   | INSERT INTO test_table_a values ( $1 )
    103 | 3335 | 23636 | UTILITY | TRUNCATE test_table_b
    103 | 3335 | 23636 | UTILITY | COMMIT
    103 | 3336 | 23636 | QUERY   | INSERT INTO test_table_b values ( $1 )
    103 | 3336 | 23636 | UTILITY | COMMIT
```

El siguiente ejemplo emite un TRUNCATE desde una llamada anidada. TRUNCATE confirma todo el trabajo realizado hasta el momento en los procedimientos internos y externos en una transacción (3344). Inicia una nueva transacción (3345). La nueva transacción se confirma cuando se produce la salida del procedimiento externo.

```
CREATE OR REPLACE PROCEDURE sp_inner(c int, d int) LANGUAGE plpgsql
AS $$
BEGIN
  INSERT INTO inner_table values (c);
  TRUNCATE outer_table;
  INSERT INTO inner_table values (d);
END;
$$;

CREATE OR REPLACE PROCEDURE sp_outer(a int, b int, c int, d int) LANGUAGE plpgsql
AS $$
BEGIN
  INSERT INTO outer_table values (a);
  Call sp_inner(c, d);
  INSERT INTO outer_table values (b);
END;
$$;

Call sp_outer(1, 2, 3, 4);

select userid, xid, pid, type, trim(text) as stmt_text
from svl_statementtext where pid = pg_backend_pid() order by xid , starttime ,
 sequence;

 userid | xid  |  pid  |  type   |
                                        stmt_text
--------+------+-------+---------
+-----------------------------------------------------------------------------------------------
    103 | 3344 | 23636 | UTILITY | Call sp_outer(1, 2, 3, 4);
    103 | 3344 | 23636 | QUERY   | INSERT INTO outer_table values ( $1 )
    103 | 3344 | 23636 | UTILITY | CALL sp_inner( $1 , $2 )
    103 | 3344 | 23636 | QUERY   | INSERT INTO inner_table values ( $1 )
```

```
    103 | 3344 | 23636 | UTILITY | TRUNCATE outer_table
    103 | 3344 | 23636 | UTILITY | COMMIT
    103 | 3345 | 23636 | QUERY   | INSERT INTO inner_table values ( $1 )
    103 | 3345 | 23636 | QUERY   | INSERT INTO outer_table values ( $1 )
    103 | 3345 | 23636 | UTILITY | COMMIT
```

El siguiente ejemplo muestra que el cursor `cur1` se cerró cuando se confirmó la instrucción TRUNCATE.

```
CREATE OR REPLACE PROCEDURE sp_open_cursor_truncate()
LANGUAGE plpgsql
AS $$
DECLARE
  rec RECORD;
  cur1 cursor for select * from test_table_a order by 1;
BEGIN
  open cur1;
  TRUNCATE table test_table_b;
  Loop
    fetch cur1 into rec;
    raise info '%', rec.c1;
    exit when not found;
  End Loop;
END
$$;

call sp_open_cursor_truncate();
ERROR: cursor "cur1" does not exist
CONTEXT: PL/pgSQL function "sp_open_cursor_truncate" line 8 at fetch
```

El siguiente ejemplo emite una instrucción TRUNCATE y no puede ser llamado desde dentro de un bloque de transacciones explícitas.

```
CREATE OR REPLACE PROCEDURE sp_truncate_atomic() LANGUAGE plpgsql
AS $$
BEGIN
  TRUNCATE test_table_b;
END;
$$;

Begin;
  Call sp_truncate_atomic();
```

```
ERROR: TRUNCATE cannot be invoked from a procedure that is executing in an atomic
 context.
HINT: Try calling the procedure as a top-level call i.e. not from within an explicit
 transaction block.
Or, if this procedure (or one of its ancestors in the call chain) was created with SET
 config options, recreate the procedure without them.
CONTEXT: SQL statement "TRUNCATE test_table_b"
PL/pgSQL function "sp_truncate_atomic" line 2 at SQL statement
```

El ejemplo siguiente muestra que un usuario que no es un superusuario ni el propietario de una tabla puede emitir una instrucción TRUNCATE en la tabla. El usuario lo hace utilizando un procedimiento almacenado de `Security Definer`. En el ejemplo se muestran las acciones siguientes:

- El usuario 1 crea la tabla `test_tbl`.
- El usuario 1 crea un procedimiento almacenado `sp_truncate_test_tbl`.
- El usuario 1 concede el privilegio EXECUTE en el procedimiento almacenado al usuario 2.
- El usuario 2 ejecuta el procedimiento almacenado para truncar la tabla `test_tbl`. En el ejemplo se muestra el recuento de filas antes y después del comando TRUNCATE.

```
set session_authorization to user1;
create table test_tbl(id int, name varchar(20));
insert into test_tbl values (1,'john'), (2, 'mary');
CREATE OR REPLACE PROCEDURE sp_truncate_test_tbl() LANGUAGE plpgsql
AS $$
DECLARE
  tbl_rows int;
BEGIN
  select count(*) into tbl_rows from test_tbl;
  RAISE INFO 'RowCount before Truncate: %', tbl_rows;
  TRUNCATE test_tbl;
  select count(*) into tbl_rows from test_tbl;
  RAISE INFO 'RowCount after Truncate: %', tbl_rows;
END;
$$ SECURITY DEFINER;
grant execute on procedure sp_truncate_test_tbl() to user2;
reset session_authorization;


set session_authorization to user2;
call sp_truncate_test_tbl();
```

```
INFO:  RowCount before Truncate: 2
INFO:  RowCount after Truncate: 0
CALL
reset session_authorization;
```

El siguiente ejemplo emite COMMIT dos veces. El primer COMMIT confirma todo el trabajo realizado en la transacción 10363 e implícitamente inicia la transacción 10364. La transacción 10364 se confirma por la segunda instrucción COMMIT.

```
CREATE OR REPLACE PROCEDURE sp_commit(a int, b int) LANGUAGE plpgsql
AS $$
BEGIN
  INSERT INTO test_table values (a);
  COMMIT;
  INSERT INTO test_table values (b);
  COMMIT;
END;
$$;

call sp_commit(1,2);

select userid, xid, pid, type, trim(text) as stmt_text
from svl_statementtext where pid = pg_backend_pid() order by xid , starttime ,
 sequence;
 userid |  xid  | pid  |  type   |
                               stmt_text
--------+-------+------+---------
+-----------------------------------------------------------------------------------------------------------
    100 | 10363 | 3089 | UTILITY | call sp_commit(1,2);
    100 | 10363 | 3089 | QUERY   | INSERT INTO test_table values ( $1 )
    100 | 10363 | 3089 | UTILITY | COMMIT
    100 | 10364 | 3089 | QUERY   | INSERT INTO test_table values ( $1 )
    100 | 10364 | 3089 | UTILITY | COMMIT
```

El siguiente ejemplo emite una instrucción ROLLBACK si `sum_vals` es mayor que 2. La primera instrucción ROLLBACK revierte todo el trabajo realizado en la transacción 10377 e inicia una nueva transacción 10378. La transacción 10378 se confirma cuando se produce la salida del procedimiento.

```
CREATE OR REPLACE PROCEDURE sp_rollback(a int, b int) LANGUAGE plpgsql
AS $$
DECLARE
  sum_vals int;
```

```
BEGIN
  INSERT INTO test_table values (a);
  SELECT sum(c1) into sum_vals from test_table;
  IF sum_vals > 2 THEN
    ROLLBACK;
  END IF;

  INSERT INTO test_table values (b);
END;
$$;

call sp_rollback(1, 2);

select userid, xid, pid, type, trim(text) as stmt_text
from svl_statementtext where pid = pg_backend_pid() order by xid , starttime ,
 sequence;

userid |  xid  | pid  |  type   |
                              stmt_text
--------+-------+------+---------
+-----------------------------------------------------------------------------------------------
    100 | 10377 | 3089 | UTILITY | call sp_rollback(1, 2);
    100 | 10377 | 3089 | QUERY   | INSERT INTO test_table values ( $1 )
    100 | 10377 | 3089 | QUERY   | SELECT sum(c1) from test_table
    100 | 10377 | 3089 | QUERY   | Undoing 1 transactions on table 133646 with current
 xid 10377 : 10377
    100 | 10378 | 3089 | QUERY   | INSERT INTO test_table values ( $1 )
    100 | 10378 | 3089 | UTILITY | COMMIT
```

## Administración de transacciones de procedimientos almacenados en modo no atómico

Un procedimiento almacenado creado en modo NONATOMIC tiene un comportamiento de control de transacciones diferente al de un procedimiento creado en modo predeterminado. De forma similar al comportamiento de confirmación automática de los comandos SQL fuera de los procedimientos almacenados, cada instrucción SQL en un procedimiento NONATOMIC se ejecuta en su propia transacción y se confirma automáticamente. Si un usuario inicia un bloque de transacción explícito en un procedimiento almacenado NONATOMIC, las instrucciones SQL del bloque no se confirman automáticamente. El bloque de transacción controla la confirmación o la reversión de las instrucciones que contiene.

En los procedimientos almacenados NONATOMIC, puede abrir un bloque de transacción explícito en el procedimiento mediante la instrucción START TRANSACTION. Sin embargo, si ya hay un bloque

de transacción abierto, esta instrucción no servirá de nada porque Amazon Redshift no admite transacciones secundarias. La transacción anterior continúa.

Cuando trabaje con bucles FOR de cursor en un procedimiento NONATOMIC, asegúrese de abrir un bloque de transacción explícito antes de iterar por los resultados de una consulta. De lo contrario, el cursor se cierra cuando la instrucción SQL dentro del bucle se confirma automáticamente.

Algunas de las consideraciones al utilizar el comportamiento del modo NONATOMIC son las siguientes:

- Cada instrucción SQL del procedimiento almacenado se confirma automáticamente si no hay ningún bloque de transacción abierto y la sesión tiene activada la opción de confirmación automática.
- Puede emitir una instrucción COMMIT/ROLLBACK/TRUNCATE para finalizar la transacción si el procedimiento almacenado se llama desde un bloque de transacción. Esto no es posible en el modo predeterminado.
- Puede emitir una instrucción START TRANSACTION para iniciar un bloque de transacción en el procedimiento almacenado.

En los siguientes ejemplos, se muestra el comportamiento de las transacciones cuando se trabaja con procedimientos almacenados NONATOMIC. La sesión de todos los ejemplos siguientes tiene activada la confirmación automática.

En el siguiente ejemplo, un procedimiento almacenado NONATOMIC tiene dos instrucciones INSERT. Cuando se llama al procedimiento fuera de un bloque de transacción, cada instrucción INSERT del procedimiento se confirma automáticamente.

```
CREATE TABLE test_table_a(v int);
CREATE TABLE test_table_b(v int);

CREATE OR REPLACE PROCEDURE sp_nonatomic_insert_table_a(a int, b int) NONATOMIC AS
$$
BEGIN
    INSERT INTO test_table_a values (a);
    INSERT INTO test_table_b values (b);
END;
$$
LANGUAGE plpgsql;
```

```
Call sp_nonatomic_insert_table_a(1,2);

Select userid, xid, pid, type, trim(text) as stmt_text
from svl_statementtext where pid = pg_backend_pid() order by xid , starttime ,
 sequence;

 userid | xid  |    pid     |  type   |               stmt_text
--------+------+------------+---------+----------------------------------------
      1 | 1792 | 1073807554 | UTILITY | Call sp_nonatomic_insert_table_a(1,2);
      1 | 1792 | 1073807554 | QUERY   | INSERT INTO test_table_a values ( $1 )
      1 | 1792 | 1073807554 | UTILITY | COMMIT
      1 | 1793 | 1073807554 | QUERY   | INSERT INTO test_table_b values ( $1 )
      1 | 1793 | 1073807554 | UTILITY | COMMIT
(5 rows)
```

No obstante, cuando el procedimiento se llama desde un bloque BEGIN..COMMIT, todas las instrucciones forman parte de la misma transacción (xid=1799).

```
Begin;
  INSERT INTO test_table_a values (10);
  Call sp_nonatomic_insert_table_a(20,30);
  INSERT INTO test_table_b values (40);
Commit;

Select userid, xid, pid, type, trim(text) as stmt_text
from svl_statementtext where pid = pg_backend_pid() order by xid , starttime ,
 sequence;

 userid | xid  |    pid     |  type   |                stmt_text
--------+------+------------+---------+-----------------------------------------
      1 | 1799 | 1073914035 | UTILITY | Begin;
      1 | 1799 | 1073914035 | QUERY   | INSERT INTO test_table_a values (10);
      1 | 1799 | 1073914035 | UTILITY | Call sp_nonatomic_insert_table_a(20,30);
      1 | 1799 | 1073914035 | QUERY   | INSERT INTO test_table_a values ( $1 )
      1 | 1799 | 1073914035 | QUERY   | INSERT INTO test_table_b values ( $1 )
      1 | 1799 | 1073914035 | QUERY   | INSERT INTO test_table_b values (40);
      1 | 1799 | 1073914035 | UTILITY | COMMIT
(7 rows)
```

En este ejemplo, hay dos instrucciones INSERT entre START TRANSACTION...COMMIT. Cuando se llama al procedimiento fuera de un bloque de transacción, las dos instrucciones INSERT se encuentran en la misma transacción (xid=1866).

```
CREATE OR REPLACE PROCEDURE sp_nonatomic_txn_block(a int, b int) NONATOMIC AS
$$
BEGIN
    START TRANSACTION;
    INSERT INTO test_table_a values (a);
    INSERT INTO test_table_b values (b);
    COMMIT;
END;
$$
LANGUAGE plpgsql;

Call sp_nonatomic_txn_block(1,2);

Select userid, xid, pid, type, trim(text) as stmt_text
from svl_statementtext where pid = pg_backend_pid() order by xid , starttime ,
 sequence;

 userid | xid  |    pid     |  type   |               stmt_text
--------+------+------------+---------+----------------------------------------
      1 | 1865 | 1073823998 | UTILITY | Call sp_nonatomic_txn_block(1,2);
      1 | 1866 | 1073823998 | QUERY   | INSERT INTO test_table_a values ( $1 )
      1 | 1866 | 1073823998 | QUERY   | INSERT INTO test_table_b values ( $1 )
      1 | 1866 | 1073823998 | UTILITY | COMMIT
(4 rows)
```

Cuando se llama al procedimiento desde dentro de un bloque BEGIN...COMMIT, la instrucción START TRANSACTION en el procedimiento no hace nada porque ya hay una transacción abierta. La instrucción COMMIT en el procedimiento confirma la transacción actual (xid=1876) e inicia una nueva.

```
Begin;
  INSERT INTO test_table_a values (10);
  Call sp_nonatomic_txn_block(20,30);
  INSERT INTO test_table_b values (40);
Commit;

Select userid, xid, pid, type, trim(text) as stmt_text
from svl_statementtext where pid = pg_backend_pid() order by xid , starttime ,
 sequence;

 userid | xid  |    pid     |  type   |               stmt_text
--------+------+------------+---------+----------------------------------------
```

```
      1 | 1876 | 1073832133 | UTILITY | Begin;
      1 | 1876 | 1073832133 | QUERY   | INSERT INTO test_table_a values (10);
      1 | 1876 | 1073832133 | UTILITY | Call sp_nonatomic_txn_block(20,30);
      1 | 1876 | 1073832133 | QUERY   | INSERT INTO test_table_a values ( $1 )
      1 | 1876 | 1073832133 | QUERY   | INSERT INTO test_table_b values ( $1 )
      1 | 1876 | 1073832133 | UTILITY | COMMIT
      1 | 1878 | 1073832133 | QUERY   | INSERT INTO test_table_b values (40);
      1 | 1878 | 1073832133 | UTILITY | COMMIT
(8 rows)
```

En este ejemplo, se muestra cómo trabajar con bucles de cursor. La tabla test_table_a tiene tres valores. El objetivo es iterar por los tres valores e insertarlos en la tabla test_table_b. Si se crea un procedimiento almacenado NONATOMIC de la siguiente forma, se generará el error de que el cursor “cur1” no existe después de ejecutar la instrucción INSERT en el primer bucle. Esto se debe a que la confirmación automática de INSERT cierra el cursor abierto.

```
insert into test_table_a values (1), (2), (3);

CREATE OR REPLACE PROCEDURE sp_nonatomic_cursor() NONATOMIC
LANGUAGE plpgsql
AS $$
DECLARE
  rec RECORD;
  cur1 cursor for select * from test_table_a order by 1;
BEGIN
  open cur1;
  Loop
    fetch cur1 into rec;
    exit when not found;
    raise info '%', rec.v;
    insert into test_table_b values (rec.v);
  End Loop;
END
$$;

CALL sp_nonatomic_cursor();

INFO:  1
ERROR:  cursor "cur1" does not exist
CONTEXT:  PL/pgSQL function "sp_nonatomic_cursor" line 7 at fetch
```

Para que el bucle de cursor funcione, colóquelo entre START TRANSACTION...COMMIT.

```
insert into test_table_a values (1), (2), (3);

CREATE OR REPLACE PROCEDURE sp_nonatomic_cursor() NONATOMIC
LANGUAGE plpgsql
AS $$
DECLARE
  rec RECORD;
  cur1 cursor for select * from test_table_a order by 1;
BEGIN
  START TRANSACTION;
  open cur1;
  Loop
    fetch cur1 into rec;
    exit when not found;
    raise info '%', rec.v;
    insert into test_table_b values (rec.v);
  End Loop;
  COMMIT;
END
$$;

CALL sp_nonatomic_cursor();

INFO:  1
INFO:  2
INFO:  3
CALL
```

## Errores de captura

En este tema se describe cómo gestiona Amazon Redshift los errores.

Cuando una consulta o un comando de un procedimiento almacenado provoca un error, las consultas posteriores no se ejecutan y la transacción se revierte. Puede administrar errores mediante un bloque EXCEPTION.

> **Note**
>
> El comportamiento predeterminado es que un error provocará que las consultas posteriores no se ejecuten, aunque no haya condiciones adicionales que generen errores en el procedimiento almacenado.

```
[ <<label>> ]
[ DECLARE
  declarations ]
BEGIN
  statements
EXCEPTION
  WHEN OTHERS THEN
    statements
END;
```

Cuando se produce una excepción y agrega un bloque de manejo de excepciones, puede escribir instrucciones RAISE y la mayoría de las demás instrucciones PL/PgSQL. Por ejemplo, puede generar una excepción con un mensaje personalizado o insertar un registro en una tabla de registro.

Al ingresar el bloque de manejo de excepciones, la transacción actual se revierte y se crea una nueva transacción para ejecutar las instrucciones en el bloque. Si las instrucciones en el bloque se ejecutan sin errores, la transacción se confirma y se vuelve a lanzar la excepción. Por último, se cierra el procedimiento almacenado.

La única condición admitida en un bloque de excepciones es OTHERS, que se asocia a cada tipo de error excepto la cancelación de la consulta. Además, si se produce un error en un bloque de manejo de excepciones, este puede ser capturado por un bloque de manejo de excepciones externo.

Cuando se produce un error en el procedimiento NONATOMIC, el error no se vuelve a generar si se gestiona mediante un bloque de excepciones. Consulte la instrucción `RAISE` de PL/pgSQL para generar una excepción detectada por el bloque de gestión de excepciones. Esta instrucción solo es válida en bloques de gestión de excepciones. Para obtener más información, consulte RAISE.

Controlar lo que ocurre después de un error en un procedimiento almacenado con el controlador CONTINUE

El controlador `CONTINUE` es un tipo de controlador de excepciones que supervisa el flujo de ejecución dentro de un procedimiento almacenado NONATOMIC. Al usarlo, puede atrapar y administrar excepciones sin finalizar el bloque de instrucciones existente. Normalmente, cuando se produce un error en un procedimiento almacenado, el flujo se interrumpe y el error se devuelve a quien llama. Sin embargo, en algunos casos prácticos, la condición de error no es lo suficientemente grave como para justificar la interrupción del flujo. Se recomienda gestionar el error de forma adecuada, utilizando la lógica de gestión de errores que elija en una transacción independiente y, a continuación, seguir ejecutando instrucciones que sigan al error. A continuación se muestra la sintaxis.

```
[ DECLARE
  declarations ]
BEGIN
  statements
EXCEPTION
  [ CONTINUE_HANDLER | EXIT_HANDLER ] WHEN OTHERS THEN
    handler_statements
END;
```

Existen varias tablas del sistema que le ayudarán a recopilar información sobre los distintos tipos de errores. Para obtener más información, consulte STL_LOAD_ERRORS, STL_ERROR y SYS_STREAM_SCAN_ERRORS. También hay tablas adicionales del sistema que puede utilizar para solucionar errores. Encontrará más información al respecto en Referencia de las tablas y vistas de sistema.

## Ejemplo

En el siguiente ejemplo se muestra cómo escribir instrucciones en el bloque de manejo de excepciones. El procedimiento almacenado utiliza el comportamiento de administración de transacciones predeterminado.

```
CREATE TABLE employee (firstname varchar, lastname varchar);
INSERT INTO employee VALUES ('Tomas','Smith');
CREATE TABLE employee_error_log (message varchar);

CREATE OR REPLACE PROCEDURE update_employee_sp() AS
$$
BEGIN
    UPDATE employee SET firstname = 'Adam' WHERE lastname = 'Smith';
    EXECUTE 'select invalid';
EXCEPTION WHEN OTHERS THEN
    RAISE INFO 'An exception occurred.';
    INSERT INTO employee_error_log VALUES ('Error message: ' || SQLERRM);
END;
$$
LANGUAGE plpgsql;

CALL update_employee_sp();

INFO:  An exception occurred.
ERROR:  column "invalid" does not exist
CONTEXT:  SQL statement "select invalid"
```

```
PL/pgSQL function "update_employee_sp" line 3 at execute statement
```

En este ejemplo, si llamamos a `update_employee_sp`, el mensaje informativo Se ha producido una excepción. se genera y el mensaje de error se inserta en la tabla de registro del registro `employee_error_log`. La excepción original se vuelve a lanzar antes de que finalice el procedimiento almacenado. En las consultas siguientes, se muestran los registros resultantes de la ejecución del ejemplo.

```
SELECT * from employee;

firstname | lastname
-----------+----------
 Tomas     | Smith

SELECT * from employee_error_log;

          message
-----------------------------------------------
 Error message: column "invalid" does not exist
```

Para obtener más información acerca de RAISE, incluida la ayuda sobre el formato, junto con una lista de niveles adicionales, consulte Instrucciones de PL/pgSQL admitidas.

En el siguiente ejemplo se muestra cómo escribir instrucciones en el bloque de manejo de excepciones. El procedimiento almacenado utiliza el comportamiento de administración de transacciones NONATOMIC. En este ejemplo, no se devuelve ningún error al autor de la llamada una vez finalizada la llamada al procedimiento. La instrucción UPDATE no se revierte debido al error en la instrucción siguiente. Se genera un mensaje informativo y se inserta el mensaje de error en la tabla de registro.

```
CREATE TABLE employee (firstname varchar, lastname varchar);
INSERT INTO employee VALUES ('Tomas','Smith');
CREATE TABLE employee_error_log (message varchar);

-- Create the SP in NONATOMIC mode
CREATE OR REPLACE PROCEDURE update_employee_sp_2() NONATOMIC AS
$$
BEGIN
    UPDATE employee SET firstname = 'Adam' WHERE lastname = 'Smith';
    EXECUTE 'select invalid';
EXCEPTION WHEN OTHERS THEN
```

```
    RAISE INFO 'An exception occurred.';
    INSERT INTO employee_error_log VALUES ('Error message: ' || SQLERRM);
END;
$$
LANGUAGE plpgsql;

CALL update_employee_sp_2();
INFO:  An exception occurred.
CALL

SELECT * from employee;

 firstname | lastname
-----------+----------
 Adam      | Smith
(1 row)

SELECT * from employee_error_log;

                    message
-----------------------------------------------
 Error message: column "invalid" does not exist
(1 row)
```

En este ejemplo, se muestra cómo crear un procedimiento con dos bloques secundarios. Cuando se llama al procedimiento almacenado, el error del primer bloque secundario se gestiona mediante su bloque de gestión de excepciones. Una vez completado el primer bloque secundario, el procedimiento continúa con la ejecución del segundo bloque secundario. Puede ver en el resultado que no se produce ningún error cuando se completa la llamada al procedimiento. Se confirman las operaciones UPDATE e INSERT en la tabla employee. Los mensajes de error de ambos bloques de excepciones se insertan en la tabla de registro.

```
CREATE TABLE employee (firstname varchar, lastname varchar);
INSERT INTO employee VALUES ('Tomas','Smith');
CREATE TABLE employee_error_log (message varchar);

CREATE OR REPLACE PROCEDURE update_employee_sp_3() NONATOMIC AS
$$
BEGIN
    BEGIN
        UPDATE employee SET firstname = 'Adam' WHERE lastname = 'Smith';
        EXECUTE 'select invalid1';
```

```
    EXCEPTION WHEN OTHERS THEN
        RAISE INFO 'An exception occurred in the first block.';
        INSERT INTO employee_error_log VALUES ('Error message: ' || SQLERRM);
    END;
    BEGIN
        INSERT INTO employee VALUES ('Edie','Robertson');
        EXECUTE 'select invalid2';
    EXCEPTION WHEN OTHERS THEN
        RAISE INFO 'An exception occurred in the second block.';
        INSERT INTO employee_error_log VALUES ('Error message: ' || SQLERRM);
    END;
END;
$$
LANGUAGE plpgsql;

CALL update_employee_sp_3();
INFO:  An exception occurred in the first block.
INFO:  An exception occurred in the second block.
CALL

SELECT * from employee;

 firstname | lastname
-----------+-----------
 Adam      | Smith
 Edie      | Robertson
(2 rows)

SELECT * from employee_error_log;

                    message
-----------------------------------------------
 Error message: column "invalid1" does not exist
 Error message: column "invalid2" does not exist
(2 rows)
```

El siguiente ejemplo muestra cómo utilizar el controlador de excepciones CONTINUE. En este ejemplo se crean dos tablas y se utilizan en un procedimiento almacenado. El controlador CONTINUE controla el flujo de ejecución en un procedimiento almacenado con un comportamiento de administración de transacciones NONATOMIC.

```
CREATE TABLE tbl_1 (a int);
CREATE TABLE tbl_error_logging(info varchar, err_state varchar, err_msg varchar);
```

```
CREATE OR REPLACE PROCEDURE sp_exc_handling_1() NONATOMIC AS
$$
BEGIN
    INSERT INTO tbl_1 VALUES (1);
    -- Expect an error for the insert statement following, because of the invalid value
    INSERT INTO tbl_1 VALUES ("val");
    INSERT INTO tbl_1 VALUES (2);
EXCEPTION CONTINUE_HANDLER WHEN OTHERS THEN
    INSERT INTO tbl_error_logging VALUES ('Encountered error', SQLSTATE, SQLERRM);
END;
$$ LANGUAGE plpgsql;
```

LLame al procedimiento almacenado:

```
CALL sp_exc_handling_1();
```

El flujo actua de la siguiente manera:

1. Se produce un error porque se intenta insertar un tipo de datos incompatible en una columna. El control pasa al bloque EXCEPTION. Al entrar en el bloque de control de excepciones, la transacción actual se revierte y se crea una nueva transacción implícita para ejecutar las instrucciones en ella.
2. Si las instrucciones de CONTINUE_HANDLER se ejecutan sin errores, el control pasa a la instrucción inmediatamente posterior a la instrucción que provoca la excepción. (Si una instrucción de CONTINUE_HANDLER genera una nueva excepción, puede gestionarla con un controlador de excepciones dentro del bloque EXCEPTION).

Después de llamar al procedimiento almacenado de ejemplo, las tablas contienen los siguientes registros:

- Si ejecuta `SELECT * FROM tbl_1;`, se devuelven dos registros que contienen los valores 1 y 2.
- Si ejecuta `SELECT * FROM tbl_error_logging;`, se devuelve un registro con los siguientes valores: Encountered error, 42703 y column "val" does not exist in tbl_1.

El siguiente ejemplo de control de errores adicional utiliza un controlador EXIT y un controlador CONTINUE. Este crea dos tablas: una tabla de datos y una tabla de registro. También crea un procedimiento almacenado que demuestra el control de errores:

```
CREATE TABLE tbl_1 (a int);
CREATE TABLE tbl_error_logging(info varchar, err_state varchar, err_msg varchar);

CREATE OR REPLACE PROCEDURE sp_exc_handling_2() NONATOMIC AS
$$
BEGIN
    INSERT INTO tbl_1 VALUES (1);
    BEGIN
        INSERT INTO tbl_1 VALUES (100);
        -- Expect an error for the insert statement following, because of the invalid
 value
        INSERT INTO tbl_1 VALUES ("val");
        INSERT INTO tbl_1 VALUES (101);
    EXCEPTION EXIT_HANDLER WHEN OTHERS THEN
        INSERT INTO tbl_error_logging VALUES ('Encountered error', SQLSTATE, SQLERRM);
    END;
    INSERT INTO tbl_1 VALUES (2);
    -- Expect an error for the insert statement following, because of the invalid value
    INSERT INTO tbl_1 VALUES ("val");
    INSERT INTO tbl_1 VALUES (3);
EXCEPTION CONTINUE_HANDLER WHEN OTHERS THEN
    INSERT INTO tbl_error_logging VALUES ('Encountered error', SQLSTATE, SQLERRM);
END;
$$ LANGUAGE plpgsql;
```

Tras crear el procedimiento almacenado, llámelo de la siguiente manera:

```
CALL sp_exc_handling_2();
```

Cuando se produce un error en el bloque de excepciones interno, que está entre corchetes entre los conjuntos internos BEGIN y END, este lo controla el controlador EXIT. Cualquier error que se produzca en el bloque exterior lo gestiona el controlador CONTINUE.

Después de llamar al procedimiento almacenado de ejemplo, las tablas contienen los siguientes registros:

- Si ejecuta `SELECT * FROM tbl_1;`, se devuelven cuatro registros con los valores 1, 2, 3 y 100.
- Si ejecuta `SELECT * FROM tbl_error_logging;`, se devuelven dos registros Estos tienen los siguientes valores: Encountered error, 42703 y column "val" does not exist in tbl_1.

Si la tabla tbl_error_logging no existe, se genera una excepción.

El siguiente ejemplo muestra cómo utilizar el controlador de excepciones CONTINUE con el bucle FOR. En este ejemplo se crean tres tablas y se utilizan en un bucle FOR dentro de un procedimiento almacenado. El bucle FOR es una variante del conjunto de resultados, lo que significa que itera sobre los resultados de una consulta:

```
CREATE TABLE tbl_1 (a int);
INSERT INTO tbl_1 VALUES (1), (2), (3);
CREATE TABLE tbl_2 (a int);
CREATE TABLE tbl_error_logging(info varchar, err_state varchar, err_msg varchar);

CREATE OR REPLACE PROCEDURE sp_exc_handling_loop() NONATOMIC AS
$$
DECLARE
 rec RECORD;
BEGIN
    FOR rec IN SELECT a FROM tbl_1
    LOOP
        IF rec.a = 2 THEN
            -- Expect an error for the insert statement following, because of the
 invalid value
            INSERT INTO tbl_2 VALUES("val");
        ELSE
            INSERT INTO tbl_2 VALUES (rec.a);
        END IF;
    END LOOP;
EXCEPTION CONTINUE_HANDLER WHEN OTHERS THEN
    INSERT INTO tbl_error_logging VALUES ('Encountered error', SQLSTATE, SQLERRM);
END;
$$ LANGUAGE plpgsql;
```

LLame al procedimiento almacenado:

```
CALL sp_exc_handling_loop();
```

Después de llamar al procedimiento almacenado de ejemplo, las tablas contienen los siguientes registros:

- Si ejecuta `SELECT * FROM tbl_2;`, se devuelven dos registros que contienen los valores 1 y 3.
- Si ejecuta `SELECT * FROM tbl_error_logging;`, se devuelve un registro con los siguientes valores: Encountered error, 42703 y column "val" does not exist in tbl_2.

Notas de uso sobre el controlador CONTINUE:

- Las palabras clave CONTINUE_HANDLER y EXIT_HANDLER solo se pueden usar en procedimientos almacenados NONATOMIC.
- Las palabras clave CONTINUE_HANDLER y EXIT_HANDLER son opcionales. EXIT_HANDLER es la opción predeterminada.

## Registro de procedimientos almacenados

En este tema se describen los procedimientos almacenados y las vistas que Amazon Redshift utiliza para el registro de procedimientos almacenados.

Los detalles sobre los procedimientos almacenados se registran en las siguientes vistas y tablas del sistema:

- SVL_STORED_PROC_CALL: se registran detalles acerca de la hora de inicio y finalización de la llamada al procedimiento almacenado y acerca de si la llamada finaliza antes de la conclusión. Para obtener más información, consulte SVL_STORED_PROC_CALL.
- SVL_STORED_PROC_MESSAGES: los mensajes de los procedimientos almacenados emitidos por la consulta RAISE se registran con el nivel de registro correspondiente. Para obtener más información, consulte SVL_STORED_PROC_MESSAGES.
- SVL_QLOG: se registra el ID de consulta de la llamada al procedimiento para cada consulta llamada desde un procedimiento almacenado. Para obtener más información, consulte SVL_QLOG.
- STL_UTILITYTEXT: se registran llamadas al procedimiento almacenado cuando están completadas. Para obtener más información, consulte STL_UTILITYTEXT.
- PG_PROC_INFO: esta vista de catálogo del sistema muestra información acerca de procedimientos almacenados. Para obtener más información, consulte PG_PROC_INFO.

## Limitaciones de los procedimientos almacenados

En este tema se describen las limitaciones de los procedimientos almacenados de Amazon Redshift.

Se aplican las siguientes consideraciones cuando utiliza procedimientos almacenados de Amazon Redshift.

## Diferencias entre Amazon Redshift y PostgreSQL para compatibilidad de procedimientos almacenados

Diferencias entre la compatibilidad de procedimientos almacenados de Amazon Redshift y de PostgreSQL:

- Amazon Redshift no admite sustracciones y, por lo tanto, tiene una compatibilidad limitada con bloques de manejo de excepciones.

## Consideraciones y límites

A continuación, se presentan consideraciones sobre los procedimientos almacenados en Amazon Redshift:

- El número máximo de procedimientos almacenados para una base de datos es de 10 000.
- El tamaño máximo del código fuente para un procedimiento es 2 MB.
- El número máximo de cursores explícitos e implícitos que puede abrir a la vez en una sesión de usuario es uno. Los bucles FOR que se repiten a lo largo del conjunto de resultados de una instrucción de SQL abren cursores implícitos. No se admiten cursores anidados.
- Los cursores explícitos e implícitos tienen las mismas restricciones de tamaño del conjunto de resultados que los cursores de Amazon Redshift estándar. Para obtener más información, consulte Restricciones del cursor.
- El número máximo de niveles para llamadas anidadas es 16.
- El número máximo de parámetros de procedimientos es 32 para argumentos de entrada y 32 para argumentos de salida.
- El número máximo de variables en un procedimiento almacenado es 1024.
- Los comandos de SQL que exigen su propio contexto de transacción no se admiten dentro de un procedimiento almacenado. Entre los ejemplos se incluyen:
  - PREPARE
  - CREATE/DROP DATABASE
  - CREATE EXTERNAL TABLE
  - VACUUM
  - SET LOCAL
  - ALTER TABLE APPEND

- La llamada del método `registerOutParameter` a través del controlador de conectividad de bases de datos de Java (JDBC) no es compatible con el tipo de datos `refcursor`. Para ver un ejemplo de uso del tipo de datos `refcursor`, consulte Devolución de un conjunto de resultados desde un procedimiento almacenado.

# Referencia del lenguaje PL/pgSQL

Los procedimientos almacenados en Amazon Redshift se basan en el lenguaje de procedimientos PL/pgSQL de PostgreSQL, con algunas diferencias importantes. En esta referencia, encontrará información detallada de la sintaxis de PL/pgSQL tal y como la implementa Amazon Redshift. Para obtener más información sobre PL/pgSQL, consulte PL/pgSQL - SQL procedural language en la documentación de PostgreSQL.

Temas


- Convenciones de la referencia de PL/pgSQL
- Estructura de PL/pgSQL
- Instrucciones de PL/pgSQL admitidas


## Convenciones de la referencia de PL/pgSQL

En esta sección se explican las convenciones que se utilizan para escribir la sintaxis del lenguaje de procedimientos almacenados PL/pgSQL.

| Carácter | Descripción |
|---|---|
| CAPS | Las palabras en mayúscula son palabras clave. |
| [ ] | Los corchetes denotan argumentos opcionales. Varios argumentos entre corchetes indican que puede seleccionar cualquier cantidad de argumento s. Además, los argumentos entre corchetes en líneas distintas indican que el analizador de Amazon Redshift espera que los argumentos estén en el orden en que aparecen en la sintaxis. |
| { } | Las llaves indican que debe seleccionar uno de los argumentos contenidos en las llaves. |

| Carácter | Descripción |
| --- | --- |
| \| | Las barras verticales indican que puede seleccionar entre los argumentos. |
| *cursiva roja* | Las palabras en cursiva roja indican marcadores de posición. Debe insertar el valor adecuado en lugar de la palabra en cursiva roja. |
| . . . | Los puntos suspensivos indican que puede repetir el elemento anterior. |
| ' | Las palabras entre comillas simples indican que debe escribir las comillas. |

# Estructura de PL/pgSQL

PL/pgSQL es un lenguaje de procedimientos con muchas de las mismas construcciones que otros lenguajes de procedimientos.

Temas

- Bloque
- Declaración de variables
- Declaración de alias
- Variables integradas
- Tipos de registro

## Bloque

PL/pgSQL es un lenguaje estructurado en bloques. El cuerpo completo de un procedimiento se define en un bloque, que contiene declaraciones de variables e instrucciones de PL/pgSQL. Una instrucción también puede ser un bloque anidado, o subbloque.

Debe finalizar las instrucciones y declaraciones con un punto y coma. Siga la palabra clave END en un bloque o subbloque con un punto y coma. No use punto y coma después de las palabras clave DECLARE y BEGIN.

Puede escribir todas las palabras clave e identificadores con una combinación de mayúsculas y minúsculas. Los identificadores se convierten en minúsculas de forma implícita a menos que estén entrecomillados (con comillas dobles).

Un guion doble (--) inicia un comentario que se extiende hasta el final de la línea. A /* marca el inicio de un bloque de comentarios que se extiende hasta la siguiente ocurrencia de */. Los bloques de comentarios no pueden anidarse. No obstante, los comentarios con guiones dobles pueden encerrarse en un bloque de comentarios, y un guion doble puede ocultar los delimitadores de bloques de comentarios /* y */.

Cualquier instrucción en la sección de instrucciones de un bloque puede ser un subbloque. Puede utilizar subbloques con fines de agrupación lógica o para localizar variables en un pequeño grupo de instrucciones.

```
[ <<label>> ]
[ DECLARE
  declarations ]
BEGIN
  statements
END [ label ];
```

Las variables declaradas en la sección de declaraciones antes de un bloque se inicializan con sus valores predeterminados cada vez que se introduce el bloque. En otras palabras, no se inicializan solo una vez por cada llamada a la función.

A continuación se muestra un ejemplo.

```
CREATE PROCEDURE update_value() AS $$
DECLARE
  value integer := 20;
BEGIN
  RAISE NOTICE 'Value here is %', value;  -- Value here is 20
  value := 50;
  --
  -- Create a subblock
  --
  DECLARE
    value integer := 80;
  BEGIN
    RAISE NOTICE 'Value here is %', value;  -- Value here is 80
  END;

  RAISE NOTICE 'Value here is %', value;  -- Value here is 50
END;
$$ LANGUAGE plpgsql;
```

Utilice una etiqueta para identificar el bloque que se usará en una instrucción EXIT o para cualificar los nombres de las variables declaradas en el bloque.

No confunda el uso de BEGIN/END para la agrupación de instrucciones en PL/pgSQL con los comandos de la base de datos para el control de transacciones. BEGIN y END en PL/pgSQL solo son para agrupar. No inician ni finalizan una transacción.

## Declaración de variables

Declare todas las variables de un bloque, excepto las variables de bucle, en la sección DECLARE del bloque. Las variables pueden utilizar cualquier tipo de datos de Amazon Redshift válido. Para conocer los tipos de datos admitidos, consulte Tipos de datos.

Las variables de PL/pgSQL pueden ser cualquier tipo de datos admitido por Amazon Redshift, además de `RECORD` y `refcursor`. Para obtener más información acerca de `RECORD`, consulte Tipos de registro. Para obtener más información acerca de `refcursor`, consulte Cursores.

```
DECLARE
name [ CONSTANT ] type [ NOT NULL ] [ { DEFAULT | := } expression ];
```

A continuación se muestran declaraciones de variables de ejemplo.

```
customerID integer;
numberofitems numeric(6);
link varchar;
onerow RECORD;
```

La variable de bucle para la iteración de un bucle FOR a lo largo de un intervalo de valores enteros se declara automáticamente como una variable con valor entero.

La cláusula DEFAULT, si se proporciona, especifica el valor inicial asignado a la variable cuando se introduce el bloque. Si la cláusula DEFAULT no se proporciona, la variable se inicializa con el valor NULL de SQL. La opción CONSTANT impide la asignación de otros valores a la variable, de modo que su valor permanece constante durante la duración del bloque. Si se especifica NOT NULL, una asignación de un valor NULL produce un error de tiempo de ejecución. Todas las variables declaradas como NOT NULL deben tener un valor no nulo especificado de manera predeterminada.

El valor predeterminado se evalúa cada vez que se introduce el bloque. Por ejemplo, la asignación de `now()` a una variable de tipo `timestamp` provoca que la variable tenga la fecha y hora actual del momento de llamar a la función, no del momento de la compilación previa de la función.

```
quantity INTEGER DEFAULT 32;
url VARCHAR := 'http://mysite.com';
user_id CONSTANT INTEGER := 10;
```

El tipo de datos `refcursor` es el tipo de datos de variables de cursor dentro de procedimientos almacenados. Un procedimiento almacenado puede devolver un valor `refcursor`. Para obtener más información, consulte Devolución de un conjunto de resultados desde un procedimiento almacenado.

### Declaración de alias

Si la firma del procedimiento almacenado omite el nombre del argumento, puede declarar un alias para el argumento.

```
name ALIAS FOR $n;
```

### Variables integradas

Se admiten las siguientes variables integradas:

- FOUND
- SQLSTATE
- SQLERRM
- GET DIAGNOSTICS integer_var := ROW_COUNT;

FOUND es una variable especial de tipo booleano. FOUND se inicializa como false en cada llamada al procedimiento. FOUND se establece mediante los siguientes tipos de instrucciones:

- SELECT INTO

  Establece FOUND en true si devuelve una fila y en false si no devuelve ninguna.
- UPDATE, INSERT y DELETE

  Establecen FOUND en true si al menos una fila se ve afectada, en false en caso contrario.
- FETCH

  Establece FOUND en true si devuelve una fila y en false si no devuelve ninguna.
- Instrucción FOR

Establece FOUND en true si la instrucción FOR se repite una o más veces, y en false en caso contrario. Esto se aplica a las tres variantes de la instrucción FOR: bucles FOR enteros, bucles FOR de conjunto de registros y bucles FOR de conjuntos de registros dinámicos.

FOUND se establece cuando se produce la salida del bucle FOR. En el tiempo de ejecución del bucle, la instrucción FOR no modifica FOUND. No obstante, puede cambiar por la ejecución de otras instrucciones dentro del cuerpo del bucle.

A continuación se muestra un ejemplo.

```
CREATE TABLE employee(empname varchar);
CREATE OR REPLACE PROCEDURE show_found()
AS  $$
DECLARE
  myrec record;
BEGIN
  SELECT INTO myrec * FROM employee WHERE empname = 'John';
  IF NOT FOUND THEN
    RAISE EXCEPTION 'employee John not found';
  END IF;
END;
$$ LANGUAGE plpgsql;
```

En un controlador de excepciones, la variable especial SQLSTATE contiene el código de error que se corresponde a la excepción que se generó. La variable especial SQLSTATE contiene el mensaje de error asociado con la excepción. Estas variables están sin definir fuera de los controladores de excepciones y muestran un error si se utilizan.

A continuación se muestra un ejemplo.

```
CREATE OR REPLACE PROCEDURE sqlstate_sqlerrm() AS
$$
BEGIN
  UPDATE employee SET firstname = 'Adam' WHERE lastname = 'Smith';
  EXECUTE 'select invalid';
  EXCEPTION WHEN OTHERS THEN
  RAISE INFO 'error message SQLERRM %', SQLERRM;
  RAISE INFO 'error message SQLSTATE %', SQLSTATE;
END;
$$ LANGUAGE plpgsql;
```

ROW_COUNT se utiliza con el comando GET DIAGNOSTICS. Muestra el número de filas procesadas por el último comando SQL enviado al motor SQL.

A continuación se muestra un ejemplo.

```
CREATE OR REPLACE PROCEDURE sp_row_count() AS
$$
DECLARE
  integer_var int;
BEGIN
  INSERT INTO tbl_row_count VALUES(1);
  GET DIAGNOSTICS integer_var := ROW_COUNT;
  RAISE INFO 'rows inserted = %', integer_var;
END;
$$ LANGUAGE plpgsql;
```

## Tipos de registro

Un tipo RECORD no es un auténtico tipo de datos, solo un marcador de posición. Las variables de tipo registro toman la estructura actual de la fila que se les asigna durante un comando SELECT o FOR. La subestructura de una variable de registro puede cambiar cada vez que se le asigne un valor. Hasta que no se asigne un valor a una variable de registro por primera vez, esta no tendrá subestructura. Los intentos de acceder a un campo en ella producen un error de tiempo de ejecución.

```
name RECORD;
```

A continuación se muestra un ejemplo.

```
CREATE TABLE tbl_record(a int, b int);
INSERT INTO tbl_record VALUES(1, 2);
CREATE OR REPLACE PROCEDURE record_example()
LANGUAGE plpgsql
AS $$
DECLARE
  rec RECORD;
BEGIN
  FOR rec IN SELECT a FROM tbl_record
  LOOP
    RAISE INFO 'a = %', rec.a;
  END LOOP;
END;
```

```
$$;
```

## Instrucciones de PL/pgSQL admitidas

Las instrucciones PL/pgSQL aumentan los comandos SQL con construcciones de procedimiento, entre las que se incluyen expresiones condicionales y bucles, para controlar el flujo lógico. Se puede usar la mayoría de los comandos SQL, lo que incluye el lenguaje de manipulación de datos (DML), como COPY, UNLOAD e INSERT, y el lenguaje de definición de datos (DDL), como CREATE TABLE. Para ver una lista completa de comandos SQL, consulte Comandos SQL. Además, Amazon Redshift admite las siguientes instrucciones PL/pgSQL.

Temas


- Asignación
- SELECT INTO
- No-op
- SQL dinámico
- Return
- Condicionales: IF
- Condicionales: CASE
- Bucles
- Cursores
- RAISE
- Control de la transacción


### Asignación

La instrucción de asignación asigna un valor a una variable. La expresión debe devolver un valor único.

```
identifier := expression;
```

También es posible usar = no estándar para la asignación, en lugar de `:=`.

Si el tipo de datos de la expresión no coincide con el tipo de datos de la variable o si la variable tiene un tamaño o precisión conocidos, el valor del resultado se convertirá de forma implícita.

A continuación se muestran ejemplos.

```
customer_number := 20;
tip := subtotal * 0.15;
```

## SELECT INTO

La instrucción SELECT INTO asigna el resultado de varias columnas (pero solo una fila) a una variable de registro o lista de variables escalares.

```
SELECT INTO target select_expressions FROM ...;
```

En la sintaxis anterior, el *destino* puede ser una variable de registro o una lista separada por comas de variables simples y campos de registro. La lista *seleccionar_expresiones* y el resto del comando son iguales que en SQL normal.

Si se usa como *destino* una lista de variables, los valores seleccionados deben coincidir exactamente con la estructura del destino o se producirá un error de tiempo de ejecución. Cuando el destino es una variable de registro, se configura automáticamente al tipo fila de las columnas resultantes de la consulta.

La cláusula INTO puede aparece casi en cualquier parte en la instrucción SELECT. Normalmente, aparece justo después de la cláusula SELECT o justo antes de la cláusula FROM. Es decir, aparece justo antes o justo después de la lista *seleccionar_expresiones*.

Si la consulta no devuelve ninguna fila, se asignan valores NULL al *destino*. Si la consulta devuelve varias filas, se asigna la primera fila al *destino* y el resto se descartan. A menos que la instrucción contenga un comando ORDER BY, la primera fila no es determinista.

Para determinar si la asignación devuelve al menos una fila, use la variable especial FOUND.

```
SELECT INTO customer_rec * FROM cust WHERE custname = lname;
IF NOT FOUND THEN
  RAISE EXCEPTION 'employee % not found', lname;
END IF;
```

Para probar si un resultado de registro es nulo, puede usar el condicional IS NULL. No hay forma de determinar si se han descartado filas adicionales. En el siguiente ejemplo, se gestiona un caso donde no se devuelven filas.

```
CREATE OR REPLACE PROCEDURE select_into_null(return_webpage OUT varchar(256))
AS $$
DECLARE
  customer_rec RECORD;
BEGIN
  SELECT INTO customer_rec * FROM users WHERE user_id=3;
  IF customer_rec.webpage IS NULL THEN
    -- user entered no webpage, return "http://"
    return_webpage = 'http://';
  END IF;
END;
$$ LANGUAGE plpgsql;
```

## No-op

La instrucción no-op (`NULL;`) es una instrucción de marcador de posición que no hace nada. La instrucción no-op puede indicar que una ramificación de una cadena IF-THEN-ELSE está vacía.

```
NULL;
```

## SQL dinámico

Para generar comandos dinámicos que puedan implicar diferentes tablas o diferentes tipos de datos cada vez que se ejecutan desde un procedimiento almacenado de PL/pgSQL, use el comando `EXECUTE`.

```
EXECUTE command-string [ INTO target ];
```

En lo anterior, *`cadena-comando`* es una expresión que contiene una cadena (de tipo texto) que contiene el comando que se va a ejecutar. Esta valor de *`cadena-comando`* se envía al motor de SQL. No se realiza sustitución de variables de PL/pgSQL en la cadena del comando. Los valores de las variables se deben insertar en la cadena en el momento de su construcción.

> **Note**
>
> No puede utilizar las instrucciones COMMIT y ROLLBACK desde un SQL dinámico. Para más información sobre el uso de las instrucciones COMMIT y ROLLBACK en un proceso almacenado, vea Administración de transacciones.

Al trabajar con comandos dinámicos, a menudo tendrá que gestionar el escape de comillas simples. Recomendamos poner el texto fijo entrecomillas en el cuerpo de la función mediante signos de dólar. Los valores dinámicos que se van a insertar en una consulta construida requieren gestión especial porque ellos mismos pueden estar entrecomillados. En el siguiente ejemplo, se supone que hay entrecomillado con signos de dólar de la función como un todo, de modo que las comillas no tienen que ser dobles.

```
EXECUTE 'UPDATE tbl SET '
  || quote_ident(colname)
  || ' = '
  || quote_literal(newvalue)
  || ' WHERE key = '
  || quote_literal(keyvalue);
```

El ejemplo anterior muestra las funciones `quote_ident(text)` y `quote_literal(text)`. Este ejemplo pasa variables que contienen identificadores de columna y tabla a la función `quote_ident`. También pasa variables que contienen cadenas literales en el comando construido a la función `quote_literal`. Ambas funciones realizan los pasos adecuados para devolver el texto que se ingresó encerrado entre comillas dobles o simples, respectivamente, con cualquier carácter especial integrado con los identificadores de escape adecuados.

El entrecomillado con signos de dólar solo es útil para entrecomillar texto fijo. No escriba el ejemplo anterior en el formato siguiente.

```
EXECUTE 'UPDATE tbl SET '
  || quote_ident(colname)
  || ' = $$'
  || newvalue
  || '$$ WHERE key = '
  || quote_literal(keyvalue);
```

Esto no se hace porque el ejemplo produce un error si el contenido de `newvalue` tiene $$. El mismo problema ocurre con otros delimitadores de dólar que pueda usar. Para entrecomillar de manera segura texto que no conoce de antemano, use la función `quote_literal`.

## Return

La instrucción RETURN vuelve al intermediario de un procedimiento almacenado.

```
RETURN;
```

A continuación se muestra un ejemplo.

```
CREATE OR REPLACE PROCEDURE return_example(a int)
AS $$
BEGIN
  FOR b in 1..10 LOOP
    IF b < a THEN
      RAISE INFO 'b = %', b;
    ELSE
      RETURN;
    END IF;
  END LOOP;
END;
$$ LANGUAGE plpgsql;
```

## Condicionales: IF

La instrucción condicional IF puede adoptar las siguientes formas en el lenguaje PL/pgSQL que utiliza Amazon Redshift:

- IF ... THEN

  ```
  IF boolean-expression THEN
    statements
  END IF;
  ```

  A continuación se muestra un ejemplo.

  ```
  IF v_user_id <> 0 THEN
    UPDATE users SET email = v_email WHERE user_id = v_user_id;
  END IF;
  ```

- IF ... THEN ... ELSE

  ```
  IF boolean-expression THEN
    statements
  ELSE
    statements
  END IF;
  ```

  A continuación se muestra un ejemplo.

```
IF parentid IS NULL OR parentid = ''
THEN
  return_name = fullname;
  RETURN;
ELSE
  return_name = hp_true_filename(parentid) || '/' || fullname;
  RETURN;
END IF;
```

- IF ... THEN ... ELSIF ... THEN ... ELSE

  La palabra clave ELSIF también puede deletrearse ELSEIF.

```
IF boolean-expression THEN
  statements
[ ELSIF boolean-expression THEN
  statements
[ ELSIF boolean-expression THEN
  statements
    ...] ]
[ ELSE
  statements ]
END IF;
```

  A continuación se muestra un ejemplo.

```
IF number = 0 THEN
  result := 'zero';
ELSIF number > 0 THEN
  result := 'positive';
ELSIF number < 0 THEN
  result := 'negative';
ELSE
  -- the only other possibility is that number is null
  result := 'NULL';
END IF;
```

## Condicionales: CASE

La instrucción condicional CASE puede adoptar las siguientes formas en el lenguaje PL/pgSQL que utiliza Amazon Redshift:

- CASE simple

```
CASE search-expression
WHEN expression [, expression [ ... ]] THEN
  statements
[ WHEN expression [, expression [ ... ]] THEN
  statements
  ... ]
[ ELSE
  statements ]
END CASE;
```

La instrucción CASE simple proporciona ejecución condicional basada en la igualdad de los operandos.

El valor de *expresión-búsqueda* se evalúa una vez y se compara sucesivamente a cada *expresión* en la cláusula WHEN. Si se encuentra una coincidencia, se ejecuta la *instrucción* correspondiente y, a continuación, el control pasa a la siguiente instrucción después de END CASE. Las expresiones WHEN posteriores no se evalúan. Si no se encuentra una coincidencia, se ejecuta la *instrucción* ELSE. No obstante, si ELSE no está presente, entonces se genera una excepción CASE_NOT_FOUND.

A continuación se muestra un ejemplo.

```
CASE x
WHEN 1, 2 THEN
  msg := 'one or two';
ELSE
  msg := 'other value than one or two';
END CASE;
```

- CASE de búsqueda

```
CASE
WHEN boolean-expression THEN
  statements
```

```
[ WHEN boolean-expression THEN
  statements
  ... ]
[ ELSE
  statements ]
END CASE;
```

La forma de búsqueda de CASE proporciona ejecución condicional basada en expresiones boolenas verdaderas.

Cada *expresión-boleana* de la cláusula WHEN se evalúa por turno, hasta encontrar una que dé como resultado true. A continuación se ejecuta la instrucción correspondiente y el control pasa a la siguiente instrucción después de END CASE. Las *expresiones* WHEN posteriores no se evalúan. Si no se encuentra ningún resultado true, se ejecutan las *instrucciones* ELSE. No obstante, si ELSE no está presente, entonces se genera una excepción CASE_NOT_FOUND.

A continuación se muestra un ejemplo.

```
CASE
WHEN x BETWEEN 0 AND 10 THEN
  msg := 'value is between zero and ten';
WHEN x BETWEEN 11 AND 20 THEN
  msg := 'value is between eleven and twenty';
END CASE;
```

## Bucles

Las instrucciones bucle pueden adoptar las siguientes formas en el lenguaje PL/pgSQL que utiliza Amazon Redshift:

- Bucle simple

```
[<<label>>]
LOOP
  statements
END LOOP [ label ];
```

Un bucle simple define un bucle incondicional que se repite indefinidamente hasta que es terminado por una instrucción EXIT o RETURN. Las instrucciones EXIT y CONTINUE pueden

utilizar la etiqueta opcional dentro de los bucles anidados para especificar a qué bucle se refiere la instrucción EXIT y CONTINUE.

A continuación se muestra un ejemplo.

```
CREATE OR REPLACE PROCEDURE simple_loop()
LANGUAGE plpgsql
AS $$
BEGIN
  <<simple_while>>
  LOOP
    RAISE INFO 'I am raised once';
    EXIT simple_while;
    RAISE INFO 'I am not raised';
  END LOOP;
  RAISE INFO 'I am raised once as well';
END;
$$;
```

- Bucle Exit

```
EXIT [ label ] [ WHEN expression ];
```

Si no está presente la*etiqueta*, se termina el bucle más interno y, a continuación, se ejecuta la instrucción después de END LOOP. Si la *etiqueta* está presente, debe ser la etiqueta del nivel actual o de uno superior del bloque o bucle anidado. A continuación, el bucle o bloque mencionado se termina y el control continúa con la instrucción después del correspondiente END del bucle o bloque.

Si se especifica WHEN, solo se produce la salida del bucle si la *expresión* es verdadera. De no ser así, el control pasa a la instrucción después de EXIT.

Puede usar EXIT con todos los tipos de bucles; su uso no se limita a bucles incondicionales.

Cuando se usa con un bloque BEGIN, EXIT pasa el control a la siguiente instrucción después del final del bloque. Debe usarse una etiqueta para este propósito. Un EXIT sin etiqueta no se considera una coincidencia para un bloque BEGIN.

A continuación se muestra un ejemplo.

```
CREATE OR REPLACE PROCEDURE simple_loop_when(x int)
```

```
LANGUAGE plpgsql
AS $$
DECLARE i INTEGER := 0;
BEGIN
  <<simple_loop_when>>
  LOOP
    RAISE INFO 'i %', i;
    i := i + 1;
    EXIT simple_loop_when WHEN (i >= x);
  END LOOP;
END;
$$;
```

- Bucle Continue

```
CONTINUE [ label ] [ WHEN expression ];
```

Si no se proporciona una *etiqueta*, la ejecución salta a la siguiente iteración del bucle más interno. Es decir, se omiten todas las instrucciones restantes en el cuerpo del bucle. El control vuelve a la expresión de control del bucle (de haberla) para determinar si se necesita otra iteración del bucle. Si la *etiqueta* está presente, especifica la etiqueta del bucle cuya ejecución se continúa.

Si se especifica WHEN, la siguiente iteración del bucle solo empieza si la *expresión* es verdadera. De no ser así, el control pasa a la instrucción después de CONTINUE.

Puede usar CONTINUE con todos los tipos de bucles; su uso no se limita a bucles incondicionales.

```
CONTINUE mylabel;
```

- Bucle WHILE

```
[<<label>>]
WHILE expression LOOP
  statements
END LOOP [ label ];
```

La instrucción WHILE repite una secuencia de instrucciones siempre que la *expresión-boleana* sea verdadera. La expresión se comprueba justo antes de cada entrada al cuerpo del bucle.

A continuación se muestra un ejemplo.

```
WHILE amount_owed > 0 AND gift_certificate_balance > 0 LOOP
  -- some computations here
END LOOP;

WHILE NOT done LOOP
  -- some computations here
END LOOP;
```

- Bucle FOR (variante de entero)

```
[<<label>>]
FOR name IN [ REVERSE ] expression .. expression LOOP
  statements
END LOOP [ label ];
```

El bucle FOR (variante de entero) crea un bucle que se repite a lo largo de un intervalo de valores enteros. El nombre de la variable se define automáticamente como de tipo entero y solo existe dentro del bucle. Se hace caso omiso de cualquier definición existente del nombre de la variable dentro del bucle. Las dos expresiones que indican los límites inferior y superior del intervalo se evalúan una vez cuando se entra en el bucle. Si especifica REVERSE, el valor de paso se sustrae, en lugar de sumarse después de cada iteración.

Si el límite inferior es mayor que el superior (o menor en el caso de REVERSE), el cuerpo del bucle no se ejecuta. No se genera un error.

Si se adjunta una etiqueta al bucle FOR, puede hacer referencia a la variable de bucle de valor entero con un nombre cualificado, usando esa etiqueta.

A continuación se muestra un ejemplo.

```
FOR i IN 1..10 LOOP
  -- i will take on the values 1,2,3,4,5,6,7,8,9,10 within the loop
END LOOP;

FOR i IN REVERSE 10..1 LOOP
  -- i will take on the values 10,9,8,7,6,5,4,3,2,1 within the loop
END LOOP;
```

- Bucle FOR (variante de conjunto de resultados)

```
[<<label>>]
FOR target IN query LOOP
  statements
END LOOP [ label ];
```

El *destino* es una variable de registro o lista separada por comas de variables escalares. Se asigna sucesivamente al destino cada fila resultante de la consulta; el cuerpo del bucle se ejecuta para cada fila.

El bucle FOR (variante de conjunto de resultados) permite que un procedimiento almacenado se repita a lo largo de los resultados de una consulta y manipula esos datos en consecuencia.

A continuación se muestra un ejemplo.

```
CREATE PROCEDURE cs_refresh_reports() AS $$
DECLARE
  reports RECORD;
BEGIN
  FOR reports IN SELECT * FROM cs_reports ORDER BY sort_key LOOP
    -- Now "reports" has one record from cs_reports
    EXECUTE 'INSERT INTO ' || quote_ident(reports.report_name) || ' ' ||
 reports.report_query;
  END LOOP;
  RETURN;
END;
$$ LANGUAGE plpgsql;
```

- Bucle FOR con SQL dinámico

```
[<<label>>]
FOR record_or_row IN EXECUTE text_expression LOOP
  statements
END LOOP;
```

El bucle FOR con SQL dinámico permite a un procedimiento almacenado repetirse a lo largo de los resultados de una consulta dinámica y manipula esos datos en consecuencia.

A continuación se muestra un ejemplo.

```
CREATE OR REPLACE PROCEDURE for_loop_dynamic_sql(x int)
```

```
LANGUAGE plpgsql
AS $$
DECLARE
  rec RECORD;
  query text;
BEGIN
  query := 'SELECT * FROM tbl_dynamic_sql LIMIT ' || x;
  FOR rec IN EXECUTE query
  LOOP
    RAISE INFO 'a %', rec.a;
  END LOOP;
END;
$$;
```

## Cursores

En lugar de ejecutar una consulta completa a la vez, puede establecer un cursor. Un cursor encapsula una consulta y después lee el resultado de la consulta una pocas filas a la vez. Una de las razones por las que se hace esto es evitar desbordar la memoria cuando el resultado contiene un gran número de filas. Otra razón es devolver una referencia a un cursor que ha creado un procedimiento almacenado, lo que permite al intermediario leer las filas. Esta estrategia ofrece una forma eficiente de devolver grandes conjuntos de filas desde los procedimientos almacenados.

Para utilizar los cursores en un procedimiento almacenado NONATOMIC, coloque el bucle de cursor entre START TRANSACTION...COMMIT.

Para establecer un cursor, primero debe declarar una variable de cursor. Todos los accesos a cursores en PL/pgSQL pasan por variables de cursor, que son siempre del tipo de datos especial `refcursor`. Un tipo de datos `refcursor` simplemente contiene una referencia a un cursor.

Una manera de crear una variable de cursor es declararla como una variable de tipo `refcursor`. También puede utilizar la siguiente sintaxis de declaración del cursor.

```
name CURSOR [ ( arguments ) ] FOR query ;
```

En lo anterior, la lista de *`argumentos`* (si se especifica) es una lista separada por comas de pares de *`nombre tipo de datos`* que definen los nombres que se sustituirán por valores de los parámetros de la *`consulta`*. Los valores reales que se sustituirán para estos nombres se especifican más adelante, cuando se abre el cursor.

A continuación se muestran ejemplos.

```
DECLARE
  curs1 refcursor;
  curs2 CURSOR FOR SELECT * FROM tenk1;
  curs3 CURSOR (key integer) IS SELECT * FROM tenk1 WHERE unique1 = key;
```

Estas tres variables tienen el tipo de datos `refcursor`, pero la primera se puede usar con cualquier consulta. En contraste, la segunda tiene una consulta totalmente especificada vinculada a ella y la última tiene una consulta parametrizada vinculada a ella. El valor `key` se sustituye por un valor de parámetro de número entero cuando se abre el cursor. Se dice que la variable `curs1` es no vinculadaporque no está vinculada a ninguna consulta en particular.

Antes de usar el cursor para recuperar filas, debe abrirse el cursor. PL/pgSQL tiene tres formas de la instrucción OPEN, dos de las cuales utilizan variables de cursor no vinculadas mientras que la tercera utiliza una variable de cursor vinculada.

- Abrir para seleccionar: la variable de cursor se abre y se le entrega la consulta específica que debe ejecutar. El cursor no puede estar abierto aun. Además, tiene que haberse declarado como un cursor no vinculado (es decir, como una variable `refcursor` simple). La consulta SELECT se trata del mismo modo que otras instrucciones SELECT en PL/pgSQL.

  ```
  OPEN cursor_name FOR SELECT ...;
  ```

  A continuación se muestra un ejemplo.

  ```
  OPEN curs1 FOR SELECT * FROM foo WHERE key = mykey;
  ```

- Abrir para ejecutar: la variable de cursor se abre y se le entrega la consulta específica que debe ejecutar. El cursor no puede estar abierto aun. Además, tiene que haberse declarado como un cursor no vinculado (es decir, como una variable `refcursor` simple). La consulta se especifica como una expresión de cadena del mismo modo que el comando EXECUTE. Esta estrategia proporciona flexibilidad porque la consulta puede variar de una ejecución a otra.

  ```
  OPEN cursor_name FOR EXECUTE query_string;
  ```

  A continuación se muestra un ejemplo.

```
OPEN curs1 FOR EXECUTE 'SELECT * FROM ' || quote_ident($1);
```

- Abrir un cursor vinculado: esta forma de OPEN se usa para abrir una variable de cursor que tenía una consulta vinculada cuando se declaró. El cursor no puede estar abierto aun. Debe aparecer una lista de expresiones de argumento si, y solo si, se había declarado que el cursor tenía argumentos. Estos valores se sustituyen en la consulta.

  ```
  OPEN bound_cursor_name [ ( argument_values ) ];
  ```

  A continuación se muestra un ejemplo.

  ```
  OPEN curs2;
  OPEN curs3(42);
  ```

Después de abrir un cursor, puede trabajar con él usando las instrucciones descritas a continuación. Estas instrucciones no tienen que ocurrir necesariamente en el mismo procedimiento almacenado en el que inicialmente se abrió el cursor. Puede devolver un valor `refcursor` de una procedimiento almacenado y dejar que el intermediario opere en el cursor. Todos los portales se cierran de forma implícita al final de la transacción. Por lo tanto, un valor `refcursor` se puede utilizar para hacer referencia a un cursor abierto solo hasta el final de la transacción.

- FETCH recupera la siguiente fila del cursor en un destino. El destino puede ser una variable de fila, una variable de registro o una lista separada por comas de variables simples, como con SELECT INTO. Al igual que con SELECT INTO, puede comprobar la variable especial FOUND para ver si se ha obtenido una fila.

  ```
  FETCH cursor INTO target;
  ```

  A continuación se muestra un ejemplo.

  ```
  FETCH curs1 INTO rowvar;
  ```

- CLOSE cierra el portar subyacente de un cursor abierto. Puede usar esta instrucción para liberar recursos antes del final de la transacción. También puede usar esta instrucción para liberar la variable de cursor para poder abrirla de nuevo.

```
CLOSE cursor;
```

A continuación se muestra un ejemplo.

```
CLOSE curs1;
```

## RAISE

Utilice la instrucción `RAISE level` para comunicar mensajes y generar errores.

```
RAISE level 'format' [, variable [, ...]];
```

Los niveles posibles son NOTICE, INFO, LOG, WARNING y EXCEPTION. EXCEPTION genera un error, que normalmente cancela la transacción en curso. Los demás niveles solo generan mensajes con distintos grados de prioridad.

Dentro de la cadena de formato, % se sustituye por la siguiente representación de la cadena del argumento opcional. Escriba %% para emitir un % literal. Actualmente, los argumentos opcionales deben ser variables simples, no expresiones, y el formato debe ser una simple cadena literal.

En el siguiente ejemplo, el valor de `v_job_id` sustituye el % en la cadena.

```
RAISE NOTICE 'Calling cs_create_job(%)', v_job_id;
```

Utilice la instrucción `RAISE` para volver a generar la excepción detectada por un bloque de gestión de excepciones. Esta instrucción solo es válida en bloques de gestión de excepciones de procedimientos almacenados de modo NONATOMIC.

```
RAISE;
```

## Control de la transacción

Puede trabajar con instrucciones de control de la transacción en el lenguaje PL/pgSQL que utiliza Amazon Redshift. Para más información sobre el uso de las instrucciones COMMIT, ROLLBACK y TRUNCATE en un proceso almacenado, vea Administración de transacciones.

En los procedimientos almacenados de modo NONATOMIC, use `START TRANSACTION` para iniciar un bloque de transacción.

```
START TRANSACTION;
```

> **Note**
>
> La instrucción START TRANSACTION de PL/pgSQL difiere del comando SQL START TRANSACTION en los siguientes aspectos:
>
> - En los procedimientos almacenados, START TRANSACTION no es sinónimo de BEGIN.
> - La instrucción PL/pgSQL no admite palabras clave opcionales de nivel de aislamiento ni de permiso de acceso.

# Vistas materializadas en Amazon Redshift

En esta sección se describe cómo crear y utilizar las vistas materializadas en Amazon Redshift. Una vista materializada es un objeto de base de datos que contiene los resultados de una consulta, que se puede usar para mejorar el rendimiento y la eficiencia.

En un entorno de almacenamiento de datos, las aplicaciones a menudo deben realizar consultas complejas en tablas de gran tamaño. Un ejemplo son las instrucciones SELECT, las cuales realizan uniones y agregaciones de varias tablas que contienen miles de millones de filas. Procesar estas consultas puede ser costoso en términos de recursos del sistema y el tiempo que se tarda en calcular los resultados.

Las vistas materializadas de Amazon Redshift proporcionan una manera de abordar estos problemas. Una vista materializada contiene un conjunto de resultados computados previamente, basados en una consulta de SQL sobre una o más tablas base. Puede emitir instrucciones SELECT para consultar una vista materializada, de la misma manera que puede consultar otras tablas o vistas en la base de datos. Amazon Redshift devuelve los resultados calculados de manera anticipada de la vista materializada, sin tener que acceder a las tablas base en absoluto. Desde el punto de vista del usuario, los resultados de la consulta se devuelven mucho más rápido en comparación con la recuperación de los mismos datos de las tablas base.

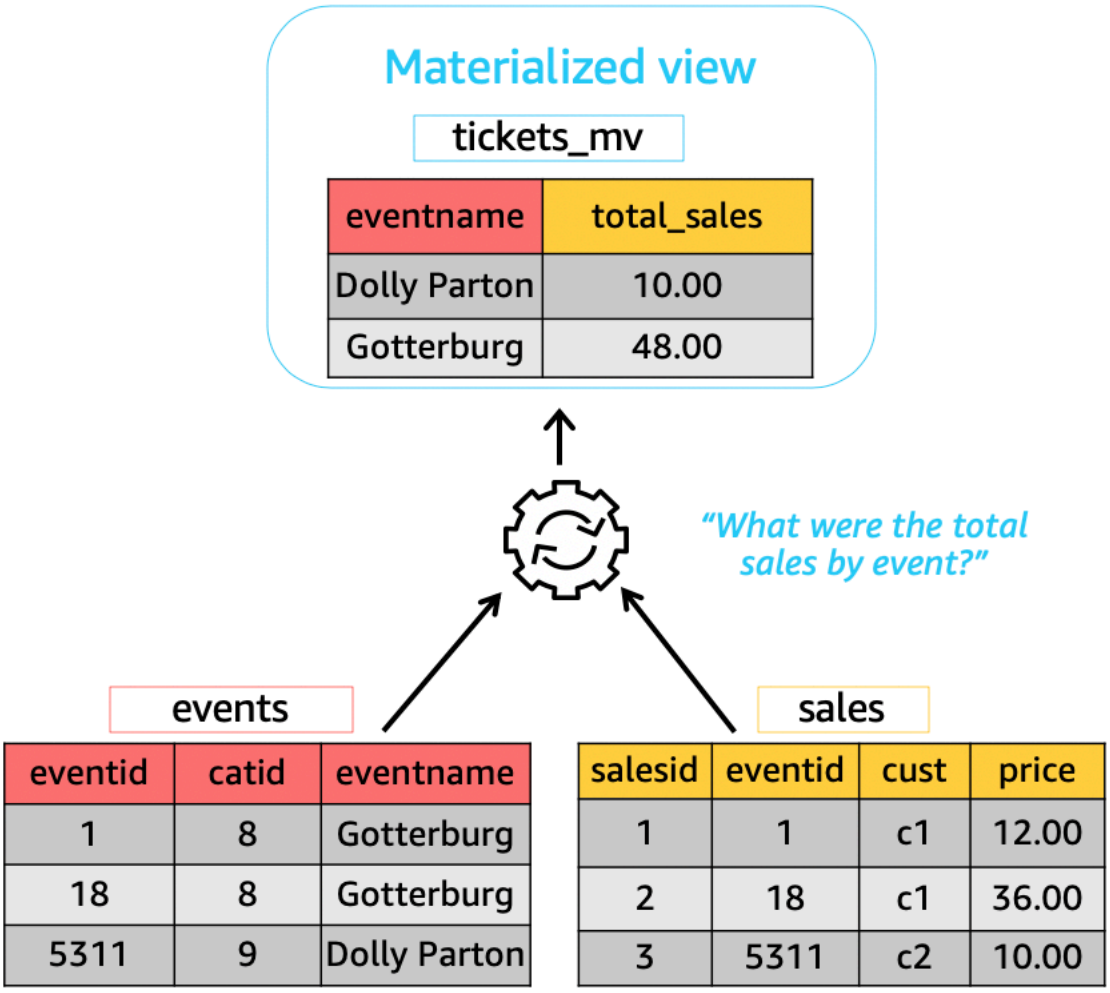
Las vistas materializadas son especialmente útiles para acelerar las consultas que son predecibles y repetidas. En lugar de realizar consultas con un uso intensivo de los recursos en grandes tablas (como las agrupaciones o las combinaciones de varios recursos), las aplicaciones pueden consultar una vista materializada y recuperar un conjunto de resultados computados con anticipación. Por ejemplo, considere el escenario en el que se utiliza un conjunto de consultas para completar paneles, como Amazon QuickSight. Este caso de uso es ideal para una vista materializada ya que las consultas son predecibles y repetidas una y otra vez.

Puede definir una vista materializada a partir de otras vistas materializadas. Use vistas materializadas en vistas materializadas para ampliar esta capacidad. En este enfoque, una vista materializada existente desempeña el mismo rol que una tabla base para que la consulta recupere datos.

Este enfoque es especialmente útil para reutilizar combinaciones calculadas con anticipación para diferentes opciones de agrupación o GROUP BY. Por ejemplo, tome una vista materializada que combine la información del cliente (que contiene millones de filas) con la información detallada de

los pedidos de elementos (que contiene miles de millones de filas). Esta es una consulta costosa de calcular bajo demanda y de forma repetida. Puede utilizar diferentes opciones GROUP BY para las vistas materializadas que se crearon encima de esta vista materializada y combinarlas con otras tablas. De este modo, se ahorra tiempo de cálculo que, de otro modo, se utilizaría para ejecutar la costosa combinación subyacente en todas las oportunidades. La tabla STV_MV_DEPS muestra las dependencias de una vista materializada en otras vistas materializadas.

Cuando crea una vista materializada, Amazon Redshift ejecuta la instrucción SQL especificada por el usuario para recopilar los datos de la tabla o las tablas base, y almacena el conjunto de resultados. En la siguiente ilustración, se proporciona información general de la vista materializada `tickets_mv` que define una consulta SQL mediante dos tablas base: `events` y `sales`.




Luego, puede utilizar estas vistas materializadas en consultas para acelerarlas. Además, Amazon Redshift puede reescribir automáticamente estas consultas para utilizar vistas materializadas, incluso cuando la consulta no referencie de manera explícita una vista materializada.

La reescritura automática de consultas es especialmente útil a la hora de mejorar el rendimiento cuando no se pueden cambiar las consultas para utilizar vistas materializadas.

Para actualizar los datos de la vista materializada, puede utilizar la instrucción REFRESH MATERIALIZED VIEW en cualquier momento para actualizarla de forma manual. Amazon Redshift identifica los cambios que han tenido lugar en la tabla o las tablas base y, luego, aplica esos cambios a la vista materializada. Dado que la reescritura automática de las consultas requiere que las vistas materializadas estén actualizadas, como propietario de vistas materializadas, asegúrese de actualizarlas cada vez que se modifique una tabla base.

Amazon Redshift proporciona algunos métodos para mantener las vistas materializadas actualizadas con objeto de llevar a cabo la reescritura automática. Puede configurar las vistas materializadas con la opción de actualización automática para actualizar las vistas materializadas cuando se actualicen sus tablas base. Esta operación de actualización automática se ejecuta en el momento en el que los recursos del clúster estén disponibles para minimizar las interrupciones en otras cargas de trabajo. Dado que la programación de la actualización automática depende de la carga de trabajo, puede tener más control sobre el momento en que Amazon Redshift actualiza las vistas materializadas. Puede programar un trabajo de actualización de la vista materializada mediante la integración de la consola y de la API del programador de Amazon Redshift. Para obtener más información acerca de la programación de consultas, visite Programar una consulta en la consola de Amazon Redshift.

Hacer esto es especialmente útil cuando existe un requisito de acuerdo de nivel de servicios (SLA) para los datos actualizados de una vista materializada. También puede actualizar de forma manual cualquier vista materializada que se pueda actualizar automáticamente. Para obtener información sobre cómo crear vistas materializadas, consulte CREATE MATERIALIZED VIEW.

Puede emitir instrucciones SELECT para consultar una vista materializada. Para obtener información sobre cómo consultar vistas materializadas, consulte Consultas de vistas materializadas. El conjunto de resultados con el tiempo se convertirá en obsoleto según se inserten, actualicen y eliminen los datos en las tablas base. Puede actualizar la vista materializada en cualquier momento para actualizarla con los últimos cambios de las tablas base. Para obtener información sobre cómo actualizar las vistas materializadas, consulte REFRESH MATERIALIZED VIEW.

Para obtener más información sobre los comandos SQL utilizados para crear y administrar vistas materializadas, consulte los siguientes temas:

- CREATE MATERIALIZED VIEW
- ALTER MATERIALIZED VIEW
- REFRESH MATERIALIZED VIEW

- DROP MATERIALIZED VIEW

Para obtener información acerca de las tablas del sistema y las vistas para supervisar las vistas materializadas, consulte los temas siguientes:

- STV_MV_INFO
- STL_MV_STATE
- SVL_MV_REFRESH_STATUS
- STV_MV_DEPS

Temas

- Consultas de vistas materializadas
- Reescritura automática de consultas para utilizar vistas materializadas
- Vistas materializadas de tablas de lagos de datos externos en Amazon Redshift Spectrum
- Actualización de una vista materializada
- Vistas materializadas automatizadas
- Uso de una función definida por el usuario (UDF) en una vista materializada
- Ingesta de streaming a una vista materializada

## Consultas de vistas materializadas

Puede utilizar una vista materializada en cualquier consulta de SQL haciendo referencia al nombre de la vista materializada como el origen de datos, como una tabla o vista estándar.

Cuando una consulta accede a una vista materializada, ve solo los datos almacenados en la vista materializada a partir de la actualización más reciente. Por lo tanto, la consulta podría no ver todos los cambios correspondientes a las tablas de base de la vista materializada.

Si otros usuarios quieren consultar la vista materializada, el propietario de la vista materializada concede el permiso SELECT a esos usuarios. Los demás usuarios no necesitan tener el permiso SELECT para las tablas base subyacentes. El propietario de la vista materializada también puede revocar el permiso SELECT a otros usuarios para impedir que consulten la vista materializada. Tenga en cuenta que los demás usuarios aún necesitan el permiso USAGE en los esquemas que contienen las tablas base de la vista materializada.

Si el propietario de la vista materializada ya no tiene el permiso SELECT local para las tablas base subyacentes:

- El propietario ya no podrá consultar la vista materializada.
- Otros usuarios que tengan el permiso SELECT para la vista materializada ya no podrán consultar la vista materializada.

Esto se limita a los permisos locales. Los cambios en los permisos administrados por Lake Formation no se verifican al consultar la vista materializada. Esto significa que si una tabla base subyacente está administrada por Lake Formation y se revocan los permisos de selección en la tabla en Lake Formation, aún puede consultar la vista materializada.

En el siguiente ejemplo, se consulta la vista materializada de `tickets_mv`. Para obtener más información sobre el comando SQL utilizado para crear una vista materializada, consulte CREATE MATERIALIZED VIEW.

```
SELECT sold
FROM tickets_mv
WHERE catgroup = 'Concerts';
```

Como los resultados de la consulta se calculan con anticipación, no es necesario acceder a las tablas subyacentes (`category`, `event` y `sales`). Amazon Redshift puede devolver los resultados directamente desde `tickets_mv`.

# Reescritura automática de consultas para utilizar vistas materializadas

Puede utilizar la reescritura automática de las consultas de vistas materializadas en Amazon Redshift para que este servicio reescriba las consultas de manera que utilicen vistas materializadas. Esto acelera las cargas de trabajo de consulta, incluso para las consultas que no referencien de manera explícita una vista materializada. Cuando Amazon Redshift reescribe consultas, solo utiliza vistas materializadas que estén actualizadas.

## Notas de uso

Para verificar si se utiliza la reescritura automática de consultas para una consulta en particular, puede inspeccionar el plan de la consulta o STL_EXPLAIN. A continuación, se muestra una instrucción SELECT y la salida EXPLAIN del plan original de la consulta.

```
SELECT catgroup, SUM(qtysold) AS sold
FROM category c, event e, sales s
WHERE c.catid = e.catid AND e.eventid = s.eventid
GROUP BY 1;

EXPLAIN
 XN HashAggregate  (cost=920021.24..920021.24 rows=1 width=35)
   ->  XN Hash Join DS_BCAST_INNER  (cost=440004.53..920021.22 rows=4 width=35)
         Hash Cond: ("outer".eventid = "inner".eventid)
         ->  XN Seq Scan on sales s  (cost=0.00..7.40 rows=740 width=6)
         ->  XN Hash  (cost=440004.52..440004.52 rows=1 width=37)
               ->  XN Hash Join DS_BCAST_INNER  (cost=0.01..440004.52 rows=1 width=37)
                     Hash Cond: ("outer".catid = "inner".catid)
                     ->  XN Seq Scan on event e  (cost=0.00..2.00 rows=200 width=6)
                     ->  XN Hash  (cost=0.01..0.01 rows=1 width=35)
                           ->  XN Seq Scan on category c  (cost=0.00..0.01 rows=1
 width=35)
```

A continuación, se muestra la salida EXPLAIN después de una reescritura automática exitosa. Esta salida incluye un análisis de la vista materializada en el plan de la consulta que reemplaza partes del plan de consulta original.

```
* EXPLAIN
     XN HashAggregate  (cost=11.85..12.35 rows=200 width=41)
       ->  XN Seq Scan on mv_tbl__tickets_mv__0 derived_table1  (cost=0.00..7.90
 rows=790 width=41)
```

Para la reescritura automática de consultas, solo se consideran las vistas materializadas actualizadas (frescas), cualquiera que sea la estrategia de actualización, como la automática, la programada o la manual. Por lo tanto, la consulta original devuelve resultados actualizados. Cuando se referencia de manera explícita una vista materializada en las consultas, Amazon Redshift accede a los datos almacenados actualmente en la vista materializada. Es posible que estos datos no reflejen los últimos cambios de las tablas base de la vista materializada.

Puede utilizar la reescritura automática de las consultas de las vistas materializadas que se crean en un clúster de la versión 1.0.20949 o versiones posteriores.

Puede detener la reescritura automática de las consultas en el nivel de la sesión configurando SET mv_enable_aqmv_for_session como FALSE.

# Cómo funciona la reescritura automática de consultas de vistas materializadas

Basándose en la optimización interna, Amazon Redshift puede decidir invocar la reescritura automática de consultas de vistas materializadas de forma transparente para proporcionar la ejecución de consultas más óptima con los tiempos de consulta más bajos posibles.

Por ejemplo, supongamos que el usuario A crea una vista materializada M1 sobre la tabla T1 con la consulta `SELECT * FROM T1`. El usuario A tiene privilegios SELECT completos en T1. El usuario A concede acceso a M1 a todos los usuarios, incluido el usuario B. No obstante, cuando el usuario B intenta consultar directamente T1, se le deniega el acceso. Esto se debe a que el usuario B no tiene privilegios SELECT en T1.

Algún tiempo después, el usuario B vuelve a intentar consultar a T1, pero esta vez obtiene resultados de T1. Esto se debe a que la reescritura automática de consultas mediante vista materializada reescribió la consulta del usuario B sobre la tabla T1 (`SELECT <cols> FROM T1`) a una consulta sobre la vista materializada M1 (`SELECT <cols> FROM M1`).

## Limitaciones

A continuación, se presentan las limitaciones para utilizar la reescritura automática de las consultas de las vistas materializadas:

- La reescritura automática de las consultas funciona con vistas materializadas que no referencian ni incluyen ninguna de las siguientes opciones:
  - subconsultas
  - combinación externa completa, izquierda o derecha
  - operaciones set
  - Cualquier función de agregación, excepto SUM, COUNT, MIN, MAX y AVG. (Estas son las únicas funciones agregadas que funcionan con la reescritura automática de consultas).
  - Cualquier función agregada con DISTINCT
  - Cualquier función de ventana
  - cláusulas SELECT DISTINCT o HAVING
  - Tablas externas o compartidas
  - otras vistas materializadas

- La reescritura automática de las consultas reescribe las consultas SELECT que referencien tablas de Amazon Redshift definidas por el usuario. Amazon Redshift no reescribe las siguientes consultas:
  - instrucciones CREATE TABLE AS
  - instrucciones SELECT INTO
  - consultas a tablas de catálogos o sistema
  - consultas con combinaciones externas o una cláusula SELECT DISTINCT
- Si una consulta no se reescribe automáticamente, verifique si tiene el permiso SELECT en la vista materializada especificada y si la opción mv_enable_aqmv_for_session está establecida en TRUE.

  También puede verificar si las vistas materializadas son aptas para la reescritura automática de las consultas inspeccionando STV_MV_INFO. Para obtener más información, consulte STV_MV_INFO.

# Vistas materializadas de tablas de lagos de datos externos en Amazon Redshift Spectrum

Las vistas materializadas pueden proporcionar un mantenimiento incremental en las tablas de lagos de datos externos. Con el mantenimiento incremental, Amazon Redshift actualiza los datos en la vista materializada únicamente con los cambios en las tablas base desde la última actualización. El mantenimiento incremental es más rentable que volver a calcular completamente la vista materializada después de cada cambio de datos en la tabla base.

Cuando se utilizan vistas materializadas en al menos una tabla externa, la creación de vistas materializadas es incremental en:

- Tablas de lagos de datos estándar, particionadas y no particionadas, con archivos de datos en cualquier formato compatible (Parquet, Avro, CSV, etc.).
- Tablas de Apache Iceberg, particionadas y no particionadas, con funciones de copiar y escribir y fusionar durante la lectura.
- Las tablas de Amazon Redshift Spectrum se unieron a cualquier tabla de Amazon Redshift de la misma base de datos.

La actualización de la vista materializada es incremental:

- Si la vista materializada no realiza agregación, se sobrescriben las tablas de lagos de datos estándar después de eliminar o PUT de S3 (eliminar archivos de datos).
- Tablas de Apache Iceberg después de INSERT, DELETE, UPDATE o la compactación de tablas.

Para obtener más información sobre Amazon Redshift Spectrum, consulte Amazon Redshift Spectrum.

## Limitaciones

Las limitaciones generales sobre las vistas materializadas siguen aplicándose a las vistas materializadas de las tablas de lagos de datos. Para obtener más información, consulte Actualización de una vista materializada. Además, tenga en cuenta las siguientes limitaciones cuando utilice vistas materializadas en tablas de lagos de datos externos.

- La creación de vistas materializadas no es incremental en:
  - Las tablas de Hudi o Delta Lake.
  - Acceso a datos anidados en Spectrum.
  - Referencias a columnas VARBYTE.
- La actualización de la vista materializada recurre al recálculo completo:
  - Apache Iceberg graba cuando una instantánea requerida ha caducado, si la vista materializada realiza la agregación.
  - Tablas de lagos de datos estándar después de eliminar o actualizar los archivos de datos en Amazon S3, si la vista materializada realiza la agregación.
  - Las tablas de lagos de datos estándar se actualizaron más de una vez dentro de un bloque de transacciones.
  - Las tablas de lagos de datos estándar se rigen por un manifiesto. Para obtener más información acerca de los manifiestos, consulte Uso de un manifiesto para especificar archivos de datos.
  - Amazon Redshift recurre al recálculo completo si se espera que tenga más rendimiento, en particular para las vistas materializadas que contienen uniones y se ha actualizado más de una tabla base desde la última actualización.
- En las tablas de Apache Iceberg, la actualización de la vista materializada solo puede gestionar hasta 4 millones de posiciones eliminadas en un único archivo de datos. Una vez alcanzado este límite, se debe compactar la tabla base de Apache Iceberg para continuar actualizando la vista materializada.

- En las tablas de Apache Iceberg, no se admite el escalado simultáneo para la creación y actualización de vistas materializadas.
- Las características autonómicas no son compatibles. Entre ellas se incluyen las vistas materializadas automatizadas, la actualización automática y la reescritura automática de consultas.
- Cuando se actualiza una vista materializada incremental, los permisos de IAM se aplican solo a las partes a las que se accede de las tablas base de Amazon Redshift.
- Los cambios en los permisos administrados por Lake Formation no se verifican al consultar una vista materializada. Esto significa que si se define una vista materializada en una tabla de lago de datos y se eliminan los privilegios de selección de la tabla con Lake Formation, aún podrá consultar la vista materializada.

# Actualización de una vista materializada

En este tema se describe cómo actualizar los datos de una vista materializada a partir de las tablas subyacentes.

Cuando crea una vista materializada, su contenido refleja el estado de la tabla o tablas de bases de datos subyacentes en ese momento. Los datos en la vista materializada permanecen sin cambios, incluso cuando las aplicaciones realizan cambios en los datos de las tablas subyacentes. Para actualizar los datos de la vista materializada, puede utilizar la instrucción `REFRESH MATERIALIZED VIEW` en cualquier momento para actualizar de forma manual las vistas materializadas. Cuando se usa esta instrucción, Amazon Redshift identifica los cambios en la tabla o las tablas base, y los aplica a la vista materializada.

Amazon Redshift tiene dos estrategias para actualizar una vista materializada:

- En muchos casos, Amazon Redshift puede realizar una actualización progresiva. En una actualización progresiva, Amazon Redshift identifica rápidamente los cambios realizados en los datos de las tablas base desde la última actualización y actualiza los datos en la vista materializada. La actualización progresiva se admite en las siguientes construcciones SQL utilizadas en la consulta a la hora de definir la vista materializada:
  - construcciones que contienen las cláusulas SELECT, FROM, [INNER] JOIN, WHERE, GROUP BY o HAVING
  - construcciones que contienen agregaciones, como SUM, MIN, MAX, AVG y COUNT
  - la mayoría de las funciones SQL integradas, específicamente las que son inmutables, dado que estas tienen los mismos argumentos de entrada y siempre producen la misma salida

La actualización incremental también es compatible con una vista materializada que se base en una tabla de recurso compartido de datos.

- Si no es posible realizar una actualización progresiva, entonces Amazon Redshift realiza una actualización completa. Una actualización completa vuelve a ejecutar la instrucción SQL subyacente, sustituyendo todos los datos en la vista materializada.
- Amazon Redshift elige automáticamente el método de actualización para una vista materializada en función de la consulta SELECT utilizada para definir la vista materializada.

Actualizar una vista materializada en una vista materializada no es un proceso en cascada. En otras palabras, supongamos que tiene una vista materializada A que depende de la vista materializada B. En este caso, cuando se invoca REFRESH MATERIALIZED VIEW A, se actualiza A utilizando la versión actual de B, incluso cuando B no está actualizada. Para que A se actualice completamente, antes de actualizar A, primero actualice B en una transacción independiente.

En el siguiente ejemplo, se muestra cómo crear un plan completo de actualización para una vista materializada mediante programación. Para actualizar la vista materializada v, primero actualice la vista materializada u. Para actualizar la vista materializada w, primero actualice la vista materializada u y, luego, la v.

```
CREATE TABLE t(a INT);
CREATE MATERIALIZED VIEW u AS SELECT * FROM t;
CREATE MATERIALIZED VIEW v AS SELECT * FROM u;
CREATE MATERIALIZED VIEW w AS SELECT * FROM v;

WITH RECURSIVE recursive_deps (mv_tgt, lvl, mv_dep) AS
( SELECT trim(name) as mv_tgt, 0 as lvl, trim(ref_name) as mv_dep
  FROM stv_mv_deps
  UNION ALL
  SELECT R.mv_tgt, R.lvl+1 as lvl, trim(S.ref_name) as mv_dep
  FROM stv_mv_deps S, recursive_deps R
  WHERE R.mv_dep = S.name
)

SELECT mv_tgt, mv_dep from recursive_deps
ORDER BY mv_tgt, lvl DESC;

 mv_tgt | mv_dep
--------+--------
 v      | u
```

```
 w      | u
 w      | v
(3 rows)
```

En el siguiente ejemplo, se muestra un mensaje informativo cuando se ejecuta REFRESH MATERIALIZED VIEW en una vista materializada que depende de una vista materializada que no está actualizada.

```
create table a(a int);
```

```
create materialized view b as select * from a;
```

```
create materialized view c as select * from b;
```

```
insert into a values (1);
```

```
refresh materialized view c;

INFO:  Materialized view c is already up to date.  However, it depends on another
 materialized view that is not up to date.
```

```
REFRESH MATERIALIZED VIEW b;
INFO:  Materialized view b was incrementally updated successfully.
```

```
REFRESH MATERIALIZED VIEW c;
INFO:  Materialized view c was incrementally updated successfully.
```

En la actualidad, Amazon Redshift tiene las siguientes limitaciones para la actualización progresiva de vistas materializadas.

Amazon Redshift no admite la actualización progresiva de las vistas materializadas que se definen con una consulta mediante los siguientes elementos SQL:

- OUTER JOIN (RIGHT, LEFT o FULL).
- las operaciones de configuración de UNION, INTERSECT, EXCEPT y MINUS

- Las funciones de agregación MEDIAN, PERCENTILE_CONT, LISTAGG, STDDEV_SAMP, STDDEV_POP, APPROXIMATE COUNT, APPROXIMATE PERCENTILE y las funciones de agregación de cada bit

  > **Note**
  >
  > Se admiten las funciones de agregación COUNT, SUM y AVG.

- Funciones de agregación DISTINCT, como DISTINCT COUNT, DISTINCT SUM, etc.
- Funciones de ventana.
- Una consulta que utiliza tablas temporales para la optimización de consultas, como la optimización de subexpresiones comunes.
- Subconsultas.
- Tablas externas que hacen referencia a los siguientes formatos de la consulta que definen la vista materializada.
  - Delta Lake
  - Hudi

  La actualización incremental es compatible con las vistas materializadas definidas con formatos distintos de los enumerados anteriormente. Para obtener más información, consulte Vistas materializadas de tablas de lagos de datos externos en Amazon Redshift Spectrum.

## Actualización automática de una vista materializada

Amazon Redshift puede actualizar automáticamente las vistas materializadas con datos actualizados de sus tablas base cuando se crean vistas materializadas con la opción de actualización automática o se modifican para activar dicha opción. Amazon Redshift actualiza automáticamente las vistas materializadas lo antes posible después de que se producen los cambios en las tablas base.

Para completar la actualización de las vistas materializadas más importantes con un impacto mínimo en las cargas de trabajo activas del clúster, Amazon Redshift tiene en cuenta varios factores. Estos factores incluyen la carga del sistema actual, los recursos necesarios para la actualización, los recursos disponibles del clúster y la frecuencia con la que se utilizan las vistas materializadas.

Amazon Redshift prioriza las cargas de trabajo antes que la actualización automática y es posible que detenga la actualización automática para preservar el rendimiento de la carga de trabajo del

usuario. Este enfoque podría retrasar la actualización de algunas vistas materializadas. En algunos casos, es posible que necesite un comportamiento de actualización más determinística para las vistas materializadas. En tal caso, considere la posibilidad de utilizar la actualización manual, como se describe en REFRESH MATERIALIZED VIEW, o la actualización programada con las operaciones de la API del programador de Amazon Redshift o la consola.

Puede establecer la actualización automática para las vistas materializadas mediante CREATE MATERIALIZED VIEW. También puede utilizar la cláusula AUTO REFRESH para actualizar las vistas materializadas de forma automática. Para obtener más información acerca de la creación de vistas materializadas, consulte CREATE MATERIALIZED VIEW. Puede activar la actualización automática de una vista materializada actual con ALTER MATERIALIZED VIEW.

Tenga en cuenta lo siguiente cuando actualice las vistas materializadas:

- Todavía puede actualizar una vista materializada de forma explícita mediante el comando REFRESH MATERIALIZED VIEW, aunque no haya habilitado la actualización automática para la vista materializada.
- Amazon Redshift no actualiza automáticamente las vistas materializadas que se definen en tablas externas.
- Para el estado de actualización, puede verificar SVL_MV_REFRESH_STATUS, que registra las consultas iniciadas por el usuario o actualizadas automáticamente.
- Para ejecutar REFRESH en vistas materializadas solo para volver a calcular, asegúrese de tener el permiso CREATE para los esquemas. Para obtener más información, consulte GRANT.

## Vistas materializadas automatizadas

En este tema se describe cómo Amazon Redshift utiliza las vistas materializadas automatizadas para mejorar el rendimiento. Amazon Redshift crea vistas materializadas de forma automática en función de la actividad y el rendimiento de la base de datos. Amazon Redshift utiliza vistas materializadas automatizadas de forma predeterminada.

Las vistas materializadas son una herramienta poderosa para mejorar el rendimiento de las consultas en Amazon Redshift. Para ello, almacenan un conjunto de resultados calculados previamente. Las consultas similares no tienen que volver a ejecutar la misma lógica cada vez, porque pueden recuperar registros del conjunto de resultados existente. Los desarrolladores y analistas crean vistas materializadas después de analizar sus cargas de trabajo para determinar qué consultas se beneficiarían y si vale la pena el costo de mantenimiento de cada vista materializada. A medida

que las cargas de trabajo crecen o cambian, estas vistas materializadas se deben revisar para asegurarse de que sigan proporcionando beneficios de rendimiento tangibles.

La característica Vista materializada automatizada (AutoMV) de Redshift supervisa las cargas de trabajo y utiliza algoritmos de machine learning para crear y gestionar vistas materializadas de manera automática, lo cual mejora el rendimiento de las consultas. Estas son algunas de las características clave de AutoMV:

- Supervisión continua: Redshift supervisa continuamente la carga de trabajo mediante técnicas de machine learning para identificar oportunidades de mejora del rendimiento mediante la creación de vistas materializadas.
- Creación y eliminación automáticas: cuando el sistema detecta que una vista materializada sería beneficiosa, la crea y mantiene automáticamente. Por el contrario, si una AutoMV creada con anterioridad deja de ofrecer mejoras de rendimiento, el sistema la eliminará automáticamente.
- No se requiere actividad del usuario: la característica AutoMV solo funciona durante los períodos de baja actividad de los usuarios o de las cargas de trabajo que se ejecutan en el clúster. Esto garantiza que las operaciones de AutoMV no interfieran ni afecten a las cargas de trabajo de los clientes.
- Aumentos de uso de la CPU: en los momentos en los que no hay actividad de la carga de trabajo, la creación o actualización de vistas materializadas por parte de AutoMV puede provocar picos de uso de la CPU. Este es un comportamiento normal, ya que el sistema utiliza los recursos disponibles para crear y actualizar las vistas materializadas.
- Prioridad de la carga de trabajo del usuario: si inicia una carga de trabajo mientras hay una operación de AutoMV en curso, la tarea de AutoMV se detendrá para liberar recursos para la carga de trabajo del usuario. Esto garantiza que sus cargas de trabajo tengan prioridad sobre las operaciones de AutoMV.

Si bien la característica AutoMV puede provocar picos de uso de la CPU durante los periodos en los que no hay actividad del usuario, funciona de forma transparente y sin afectar a las cargas de trabajo. El sistema gestiona las vistas materializadas para mejorar el rendimiento de las consultas y, al mismo tiempo, prioriza las cargas de trabajo de los usuarios por encima de las operaciones de AutoMV.

El comportamiento y las capacidades de AutoMV son los mismos que las vistas materializadas creadas por el usuario. Se actualizan de forma automática y progresiva, utilizando los mismos criterios y las mismas restricciones. Al igual que las vistas materializadas que crean los usuarios, Reescritura automática de consultas para utilizar vistas materializadas identifica las consultas que se

pueden beneficiar de los AutoMV creados por el sistema. Reescribe automáticamente esas consultas para usar AutoMVS, lo que mejora el rendimiento de las consultas. Los desarrolladores no tienen que revisar las consultas para aprovechar AutoMV.

> **Note**
>
> Las vistas materializadas automatizadas se actualizan de forma intermitente. Las consultas reescritas que utilizan AutoMV siempre devuelven los últimos resultados. Cuando Redshift detecta que los datos no están actualizados, las consultas no se reescriben para leerlas desde vistas materializadas automatizadas. En su lugar, las consultas seleccionan los datos más recientes de las tablas base.

Cualquier carga de trabajo con consultas que se utilizan de forma repetida puede aprovechar AutoMV. Los casos de uso comunes incluyen:

- Paneles: los paneles se utilizan ampliamente para proporcionar vistas rápidas de indicadores empresariales clave (KPI), eventos, tendencias y otras métricas. A menudo tienen un diseño común con gráficos y tablas, pero muestran distintas vistas para el filtrado o para las operaciones de selección de dimensiones, como desglosar. Los paneles suelen tener un conjunto común de consultas que se utilizan repetidamente con distintos parámetros. Las consultas del panel pueden beneficiarse enormemente de las vistas materializadas automatizadas.
- Informes: las consultas de informes se pueden programar en varias frecuencias, según los requisitos empresariales y el tipo de informe. Además, pueden ser automatizadas o bajo demanda. Una característica habitual de las consultas de informes es que pueden ser de ejecución prolongada y hacer un uso intensivo de recursos. Con AutoMV, no es necesario volver a calcular estas consultas cada vez que se ejecutan, lo que reduce el tiempo de ejecución de cada consulta y la utilización de recursos en Redshift.

Para desactivar las vistas materializadas automatizadas, se actualiza el grupo de parámetros `auto_mv` a `false`. Para obtener más información, consulte Grupos de parámetros de Amazon Redshift en la Guía de administración de clústeres de Amazon Redshift.

## Alcance y consideraciones de SQL de las vistas materializadas automatizadas

- Una vista materializada automatizada se puede iniciar y crear mediante una consulta o subconsulta, siempre que contenga una cláusula `GROUP BY` o una de las siguientes funciones agregadas: SUM, COUNT, MIN, MAX o AVG. Pero no puede contener ninguna de las siguientes:
  - combinación externa completa, izquierda o derecha
  - Funciones de agregación distintas a SUM, COUNT, MIN, MAX y AVG. (Estas funciones específicas funcionan con la reescritura automática de consultas).
  - Cualquier función agregada que incluya DISTINCT
  - Cualquier función de ventana
  - cláusulas SELECT DISTINCT o HAVING
  - otras vistas materializadas

  No está garantizado que una consulta que cumpla los criterios inicie la creación de una vista materializada automatizada. El sistema determina desde qué candidatos se va a crear una vista, en función de su beneficio esperado para la carga de trabajo y el costo de los recursos que se deben mantener, lo que incluye el costo de actualización del sistema. Cada vista materializada resultante se puede utilizar mediante la reescritura automática de consultas.
- Aunque AutoMV pueda iniciarse mediante una subconsulta o tramos individuales de operadores de conjunto, la vista materializada resultante no contendrá subconsultas ni operadores de conjunto.
- Para determinar si AutoMV se utilizó para consultas, consulte el plan EXPLAIN y busque `%_auto_mv_%` en la salida. Para obtener más información, consulte EXPLAIN.
- Las vistas materializadas automatizadas no son compatibles con las tablas externas, como los recursos compartidos de datos y las tablas federadas.

## Limitaciones de las vistas materializadas automatizadas

A continuación, se indican las limitaciones para trabajar con las vistas materializadas automatizadas:

- Número máximo de AutoMV: el límite de vistas materializadas automatizadas es de 200 por base de datos en el clúster.
- Espacio y capacidad de almacenamiento: una característica importante de AutoMV es que se lleva a cabo utilizando ciclos en segundo plano de reserva que contribuyen a lograr que las cargas de

trabajo de los usuarios no se vean afectadas. Si el clúster está ocupado o se está quedando sin espacio de almacenamiento, AutoMV cesa su actividad. Concretamente, al 80 % de la capacidad total del clúster no se crean nuevas vistas materializadas automatizadas. Al 90 % de la capacidad total, pueden descartarse para facilitar que las cargas de trabajo de usuario continúen sin deterioro del rendimiento. Para obtener más información sobre la determinación de la capacidad del clúster, consulte STV_NODE_STORAGE_CAPACITY.

## Facturación para vistas materializadas automatizadas

La capacidad de optimización automática de Amazon Redshift crea y actualiza vistas materializadas automatizadas. No hay ningún cargo por los recursos informáticos para este proceso. El almacenamiento de las vistas materializadas automatizadas se cobra a la tarifa normal de almacenamiento. Para obtener más información, consulte Precios de Amazon Redshift.

## Recursos adicionales

La entrada de blog que se muestra a continuación proporciona una explicación más detallada sobre las vistas materializadas automatizadas. Detalla cómo se crean, mantienen y eliminan. También explica los algoritmos subyacentes que impulsan estas decisiones: Optimize your Amazon Redshift query performance with automated materialized views (Optimice el rendimiento de las consultas de Amazon Redshift con vistas materializadas automatizadas).

Este vídeo comienza con una explicación de las vistas materializadas y muestra cómo mejoran el rendimiento y conservan los recursos. A continuación, proporciona una explicación detallada de las vistas materializadas automatizadas con una animación del flujo del proceso y una demostración en directo.

# Uso de una función definida por el usuario (UDF) en una vista materializada

Puede utilizar una UDF escalar en una vista materializada de Amazon Redshift. Defina estos valores en python o SQL y haga referencia a ellos en la definición de la vista materializada.

## Referencia a una UDF en una vista materializada

En el siguiente procedimiento, se muestra cómo utilizar las UDF que realizan comparaciones aritméticas simples, en una definición de vista materializada.

1. Cree una tabla para utilizarla en la definición de vista materializada.

```
CREATE TABLE base_table (a int, b int);
```

2. Cree una función escalar definida por el usuario en python que devuelva un valor booleano que indique si un entero es mayor que un entero de comparación.

```
CREATE OR REPLACE FUNCTION udf_python_bool(x1 int, x2 int) RETURNS bool IMMUTABLE
AS $$
  return x1 > x2
$$ LANGUAGE plpythonu;
```

   También puede crear una UDF funcionalmente similar con SQL, que puede utilizar para comparar los resultados con la primera.

```
CREATE OR REPLACE FUNCTION udf_sql_bool(int, int) RETURNS bool IMMUTABLE
AS $$
  select $1 > $2;
$$ LANGUAGE SQL;
```

3. Cree una vista materializada que seleccione de la tabla que creó y haga referencia a la UDF.

```
CREATE MATERIALIZED VIEW mv_python_udf AS SELECT udf_python_bool(a, b) AS a FROM
 base_table;
```

   También puede crear una vista materializada que haga referencia a la UDF de SQL.

```
CREATE MATERIALIZED VIEW mv_sql_udf AS SELECT udf_sql_bool(a, b) AS a FROM
 base_table;
```

4. Agregue datos a la tabla y actualice la vista materializada.

```
INSERT INTO base_table VALUES (1,2), (1,3), (4,2);
```

```
REFRESH MATERIALIZED VIEW mv_python_udf;
```

   También puede actualizar la vista materializada que hace referencia a la UDF de SQL.

```
REFRESH MATERIALIZED VIEW mv_sql_udf;
```

5. Consulte los datos desde la vista materializada.

```
SELECT * FROM mv_python_udf ORDER BY a;
```

Los resultados de la consulta son los siguientes:

```
a
-----
false
false
true
```

Devuelve `true` para el último conjunto de valores porque el valor de la columna a (4) es mayor que el valor de la columna b (2).

6. También puede consultar la vista materializada que hace referencia a la UDF de SQL. Los resultados de la función SQL coinciden con los resultados de la versión de Python.

```
SELECT * FROM mv_sql_udf ORDER BY a;
```

Los resultados de la consulta son los siguientes:

```
a
-----
false
false
true
```

Devuelve `true` para el último conjunto de valores de la comparación.

7. Utilice una instrucción DROP con CASCADE para eliminar la función definida por el usuario y la vista materializada que hace referencia a ella.

```
DROP FUNCTION udf_python_bool(int, int) CASCADE;
```

```
DROP FUNCTION udf_sql_bool(int, int) CASCADE;
```

# Ingesta de streaming a una vista materializada

En este tema se describe cómo utilizar las vistas materializadas para obtener un acceso rápido a los datos de streaming.

La ingesta de streaming proporciona una ingesta de datos de alta velocidad y baja latencia de Amazon Kinesis Data Streams o Amazon Managed Streaming para Apache Kafka a una base de datos de Amazon Redshift aprovisionada o a una base de datos de Amazon Redshift sin servidor. Los datos llegan a una vista materializada de Redshift que está configurada para el propósito. Eso da como resultado un acceso rápido a datos externos. La ingesta de streaming reduce el tiempo de acceso a los datos y reduce el costo de almacenamiento. Puede configurarlo para el clúster de Amazon Redshift o para el grupo de trabajo de Amazon Redshift sin servidor, con una pequeña recopilación de comandos SQL. Después de la configuración, cada actualización de la vista materializada, puede ingerir cientos de megabytes de datos por segundo.

## Cómo fluyen los datos de un servicio de streaming a Redshift

Ayuda a entender cómo funciona la ingesta de streaming y los objetos de base de datos que se utilizan en el proceso. Los datos fluyen directamente de un proveedor de flujo de datos a un clúster aprovisionado de Amazon Redshift o a un grupo de trabajo de Amazon Redshift sin servidor. No hay una zona de aterrizaje temporal, como un bucket de Amazon S3. El clúster aprovisionado o grupo de trabajo es el consumidor de flujos. En la base de datos de Redshift, los datos leídos del flujo aterrizan en una vista materializada. Los datos se procesan a medida que llegan. Por ejemplo, los valores JSON se pueden consumir y asignar a las columnas de datos de una vista materializada, mediante SQL. Cuando la vista materializada se actualiza, Redshift consume datos de las particiones asignadas de datos de Kinesis o particiones de Kafka hasta que la vista se actualiza con el flujo.

Los casos de uso para la ingesta de streaming de Amazon Redshift implican datos que se generan continuamente y que se deben procesar en un periodo corto o latencia, desde su generación. Esto se denomina comúnmente análisis casi en tiempo real. Los orígenes pueden incluir dispositivos TI, dispositivos de telemetría del sistema y datos de secuencias de clics de un sitio web o una aplicación muy activos.

## Prácticas recomendadas de análisis de datos para mejorar el rendimiento

Al configurar la ingesta de streaming, hay opciones sobre cómo puede analizar los datos entrantes. Las prácticas pueden incluir la lógica empresarial o el formateo a medida que llegan los datos. Sugerimos las siguientes prácticas recomendadas para ayudarle a evitar errores o la pérdida de

datos. Estas se derivan de pruebas internas y ayudan a los clientes a solucionar problemas de configuración y análisis.

- Extracción de valores de los datos transmitidos: si utiliza la función JSON_EXTRACT_PATH_TEXT en la definición de vista materializada para analizar o destruir JSON transmitido, puede afectar considerablemente al rendimiento y a la latencia. La explicación es que, por cada columna extraída con JSON_EXTRACT_PATH_TEXT, se vuelve a analizar el JSON entrante. Después de esto, se produce la conversión de tipos de datos, el filtrado y los cálculos de lógica empresarial. Esto significa, por ejemplo, que si extrae 10 columnas de los datos JSON, cada registro JSON se analiza 10 veces, lo que incluye lógica adicional. Esto se traduce en una mayor latencia de ingesta. Un enfoque alternativo que recomendamos es utilizar la función JSON_PARSE para convertir los registros JSON al tipo de datos SUPER de Redshift. Una vez que los datos transmitidos lleguen a la vista materializada, use PartiQL para extraer cadenas individuales de la representación de SUPER de los datos JSON. Para obtener más información, consulte Consulta de datos semiestructurados.

  Además, tenga en cuenta que JSON_EXTRACT_PATH_TEXT tiene un máximo de tamaño de datos de 64 KB. Por lo tanto, si algún registro JSON tiene más de 64 KB, al procesarlo con JSON_EXTRACT_PATH_TEXT se produce un error.
- Asignación de un flujo de Amazon Kinesis Data Streams o un tema de Amazon MSK a una vista materializada múltiple: no recomendamos crear varias vistas materializadas para ingerir datos desde un solo flujo o tema. Esto se debe a que cada vista materializada crea un consumidor para cada porción del flujo o la partición de Kinesis Data Streams en el tema de Kafka. Como resultado, puede producirse una limitación o superación del rendimiento del flujo o el tema. También puede suponer un costo mayor, ya que se ingieren los mismos datos varias veces. Cuando configure la ingesta de streaming, le recomendamos que cree una vista materializada para cada flujo o tema.

  Si el caso de uso requiere ingerir datos de un flujo de KDS o un tema de MSK en varias vistas materializadas, antes de hacerlo, consulte el Blog de macrodatos de AWS, específicamente la publicación Best practices to implement near-real-time analytics using Amazon Redshift Streaming Ingestion with Amazon MSK.

## Comportamiento y tipos de datos de ingesta de streaming

En la siguiente tabla se describen los detalles del comportamiento técnico y los límites de tamaño de los distintos tipos de datos. Le recomendamos que se familiarice con ellos antes de configurar una vista materializada para la ingesta de streaming.

| Característica o comportamiento | Descripción |
| --- | --- |
| Límite de longitud de los temas de Kafka | No es posible utilizar un tema de Kafka con un nombre de más de 128 caracteres (sin incluir las comillas). Para obtener más información, consulte Nombres e identificadores. |
| Actualizaciones incrementales e instrucciones JOIN en una vista materializada | La vista materializada debe ser mantenible de forma incremental. El recálculo completo no es posible para Kinesis ni Amazon MSK porque, de forma predeterminada, no conservan el historial de flujos o temas más allá de 24 horas o 7 días. Puede establecer periodos de retención de datos más largos en Kinesis o Amazon MSK. No obstante, esto puede suponer un mayor mantenimiento y costo. Además, las instrucciones JOIN no se admiten actualmente en vistas materializadas creadas en un flujo de Kinesis, o en un tema de Amazon MSK. Después de crear una vista materializada en su flujo o tema, puede crear otra para unir su vista materializada de streaming a otras vistas materializadas, tablas o vistas.<br><br>Para obtener más información, consulte REFRESH MATERIALIZED VIEW. |
| Análisis de registros | La ingesta de streaming de Amazon Redshift no admite el análisis de registros agregados por Kinesis Producer Library (Conceptos clave de KPL: agregación). Los registros agregados se ingieren, pero se almacenan como datos de búfer de protocolo binario. Para obtener más información, consulte Búferes de protocolo. Según cómo inserte los datos en Kinesis, es posible que tenga que desactivar esta característica. |
| Valores duplicados en los encabezados de Kafka | El cliente consumidor de streaming de Amazon Redshift para temas de Kafka procedentes de Amazon MSK, Confluent o Apache Kafka no admite valores duplicados en los encabezados de los temas de Kafka. |
| Descompresión | `VARBYTE` no admite la descompresión. Debido a esto, los registros que contienen datos comprimidos no se pueden |

| Característica o comportamiento | Descripción |
| --- | --- |
| | consultar en Redshift. Descomprima los datos antes de agregarlos al flujo de Kinesis o al tema de Amazon MSK. |
| Tamaño de registro máximo | El tamaño máximo de cualquier registro que Amazon Redshift puede ingerir desde Kinesis o Amazon MSK es de 16 777 216 bytes (16 MiB), el tamaño máximo que admite el tipo de datos VARBYTE en Amazon Redshift. De forma predeterminada, las vistas materializadas de streaming de Amazon Redshift creadas en un flujo de datos de Kinesis o en un tema de Amazon MSK establecerán el tamaño de la columna de datos en 1 048 576 bytes (1 MiB) y 16 777 216 bytes (16 MiB), respectivamente.<br><br>**Note**<br>1 MiB es el tamaño máximo actual de cualquier registro que se pueda incluir en un flujo de datos de Kinesis. Para obtener más información sobre los límites de tamaño de Kinesis, consulte Cuotas y límites en la Guía para desarrolladores de Amazon Kinesis Data Streams. |
| Registros de errores | En cada caso en que no se pueda realizar la ingesta de un registro en Redshift porque el tamaño de los datos supera el máximo, se omitirá ese registro. La actualización de la vista materializada se sigue realizando correctamente en este caso y se escribe un segmento de cada registro de error en la tabla del sistema SYS_STREAM_SCAN_ERRORS. Los errores que son resultado de la lógica empresarial, como un error en un cálculo o un error resultante de una conversión de tipo, no se omiten. Pruebe la lógica cuidadosamente antes de agregarla a la definición de la vista materializada. |

| Característica o comportamiento | Descripción |
| --- | --- |
| Conectividad privada de varias VPC de Amazon MSK | La conectividad privada de varias de Amazon MSK no es compatible actualmente con la ingesta de streaming de Redshift. Como alternativa, puede utilizar el emparejamiento de VPC para conectar VPC o AWS Transit Gateway para conectar VPC y redes en las instalaciones a través de un hub central. Cualquiera de estas opciones permite a Redshift comunicarse con un clúster de Amazon MSK o con Amazon MSK sin servidor en otra VPC. |
| Actualización automática, uso y activación | Las consultas de actualización automática para una vista o vistas materializadas se tratan como cualquier otra carga de trabajo de usuario. La actualización automática carga los datos del flujo a medida que llegan.<br><br>La actualización automática se puede activar explícitamente para una vista materializada creada para la ingesta de streaming. Para ello, especifique `AUTO REFRESH` en la definición de la vista materializada. La actualización manual es la opción predeterminada. Para especificar la actualización automática de una vista materializada existente para la ingesta de streaming, puede ejecutar `ALTER MATERIALIZED VIEW` para activarla. Para obtener más información, consulte CREATE MATERIALIZED VIEW o ALTER MATERIALIZED VIEW. |
| Ingesta de streaming y Amazon Redshift sin servidor | Las instrucciones de instalación y configuración que se aplican a la ingesta de streaming de Amazon Redshift en un clúster aprovisionado también se aplican a la ingesta de streaming en Amazon Redshift sin servidor. Es importante para especificar el nivel necesario de RPU para respaldar la ingesta de streaming con actualización automática y otras cargas de trabajo. Para obtener más información, consulte Facturación de Amazon Redshift sin servidor. |

| Característica o comportamiento | Descripción |
| --- | --- |
| Nodos de Amazon Redshift en una zona de disponibilidad diferente que el clúster de Amazon MSK | Cuando configure la ingesta de streaming, Amazon Redshift intenta conectarse a un clúster de Amazon MSK en la misma zona de disponibilidad, si el reconocimiento de bastidor está habilitado para Amazon MSK. Si todos los nodos se encuentran en zonas de disponibilidad distintas a la del clúster de Amazon Redshift, puede incurrir en costos de transferencia de datos entre zonas de disponibilidad. Para evitarlo, mantenga al menos un nodo del clúster de agentes de Amazon MSK en la misma zona de disponibilidad que el clúster o grupo de trabajo aprovisionados de Redshift. |
| Actualización de la ubicación de inicio | Después de crear una vista materializada, su actualización inicial comienza desde `TRIM_HORIZON` de un flujo de Kinesis o desde el desplazamiento 0 de un tema de Amazon MSK. |
| Formatos de datos | Los formatos de datos admitidos se limitan a aquellos que se pueden convertir desde `VARBYTE`. Para obtener más información, consulte Tipo VARBYTE y Operadores VARBYTE. |
| Agregación de registros a una tabla | Puede ejecutar `ALTER TABLE APPEND` para agregar filas a una tabla de destino desde una vista materializada de origen existente. Esto solo funciona si la vista materializada está configurada para la ingesta de streaming. Para obtener más información, consulte ALTER TABLE APPEND. |

| Característica o comportamiento | Descripción |
| --- | --- |
| Ejecución de TRUNCATE o DELETE | Puede eliminar registros de una vista materializada que se usa para la ingesta de streaming, mediante lo siguiente:<br>• `TRUNCATE`: esto elimina todas las filas de una vista materializada configurada para la ingesta de streaming. No hace un análisis de la tabla. Para obtener más información, consulte TRUNCATE.<br>• `DELETE`: esto elimina todas las filas de una vista materializada configurada para la ingesta de streaming. Para obtener más información, consulte DELETE. |

| Característica o comportamiento | Descripción |
| --- | --- |
| Identificadores sin minúsculas | Al crear vistas materializadas de streaming en temas de Amazon Managed Streaming para Apache Kafka o flujos de datos de Kinesis que contengan identificadores que no estén en minúsculas, es posible que se produzcan errores en las actualizaciones automáticas. Para resolver este problema, siga uno de estos pasos:<br>• Utilice actualizaciones manuales con el parámetro `enable_case_sensitive_identifier` establecido en `true`<br>• Habilite las actualizaciones automáticas mediante el establecimiento de `enable_case_sensitive_identifier` en `true` en el nivel de base de datos o clúster<br>**Note**<br>La configuración de `enable_case_sensitive_identifier` en el nivel de usuario no es suficiente para las actualizaciones automáticas, pero funcionará en las actualizaciones manuales.<br>Para obtener más información sobre los identificadores que distinguen mayúsculas de minúsculas, consulte enable_case_sensitive_identifier. |

## Introducción a la ingesta de streaming de Amazon Kinesis Data Streams

En este tema se describe cómo consumir datos de streaming de Kinesis Data Streams mediante una vista materializada.

Configurar la ingesta de streaming de Amazon Redshift implica crear un esquema externo que se asigna al origen de datos de streaming y crear una vista materializada que haga referencia a

ese esquema externo. La ingesta de streaming de Amazon Redshift admite Kinesis Data Streams como origen. Por ello, debe disponer de un origen de Kinesis Data Streams antes de configurar la ingesta de streaming. Si no dispone de un origen, siga las instrucciones de la documentación de Kinesis en Introducción a Amazon Kinesis Data Streams, o bien cree uno en la consola mediante las instrucciones de Creación de un flujo a través de la consola de administración de AWS.

La ingesta de streaming de Amazon Redshift utiliza una vista materializada, que se actualiza directamente desde el flujo cuando se ejecuta `REFRESH`. La vista materializada se asigna al origen de datos del flujo. Puede realizar filtrado y agregaciones en los datos del flujo como parte de la definición de la vista materializada. La vista materializada de ingesta de streaming (la vista materializada base) solo puede hacer referencia a un flujo, pero se pueden crear vistas materializadas adicionales que se unan con la vista materializada base y con otras tablas o vistas materializadas.

> **Note**
>
> Ingesta de streaming y Amazon Redshift sin servidor: los pasos de configuración de este tema se aplican a los clústeres de Amazon Redshift y Amazon Redshift sin servidor aprovisionados. Para obtener más información, consulte Comportamiento y tipos de datos de ingesta de streaming.

Suponiendo que disponga de un flujo de Kinesis Data Streams, el primer paso es definir un esquema en Amazon Redshift con `CREATE EXTERNAL SCHEMA` y hacer referencia a un recurso de Kinesis Data Streams. Después de eso, para acceder a los datos del flujo defina `STREAM` en una vista materializada. Puede almacenar registros del flujo con el formato `SUPER` semiestructurado, o bien definir un esquema cuyo resultado sean datos convertidos a tipos de datos de Redshift. Cuando se consulta la vista materializada, los registros devueltos son una vista de un punto temporal del flujo.

1. Cree un rol de IAM con una política de confianza que permita que el clúster de Amazon Redshift o el grupo de trabajo de Amazon Redshift sin servidor asuma el rol. Para obtener información sobre cómo configurar la política de confianza del rol de IAM, consulte Autorización a Amazon Redshift para acceder a otros servicios de AWS en su nombre. Una vez creado, el rol debe tener la siguiente política de IAM, que proporciona permiso para comunicarse con el flujo de datos de Amazon Kinesis.

   Política de IAM para un flujo no cifrado de Kinesis Data Streams

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "ReadStream",
            "Effect": "Allow",
            "Action": [
                "kinesis:DescribeStreamSummary",
                "kinesis:GetShardIterator",
                "kinesis:GetRecords",
                "kinesis:ListShards",
                "kinesis:DescribeStream"
            ],
            "Resource": "arn:aws:kinesis:*:0123456789:stream/*"
        },
        {
            "Sid": "ListStream",
            "Effect": "Allow",
            "Action": "kinesis:ListStreams",
            "Resource": "*"
        }
    ]
}
```

Política de IAM para un flujo cifrado de Kinesis Data Streams

```
{
 "Version": "2012-10-17",
 "Statement": [{
   "Sid": "ReadStream",
   "Effect": "Allow",
   "Action": [
    "kinesis:DescribeStreamSummary",
    "kinesis:GetShardIterator",
    "kinesis:GetRecords",
    "kinesis:ListShards",
    "kinesis:DescribeStream"
   ],
   "Resource": "arn:aws:kinesis:*:0123456789:stream/*"
  },
  {
   "Sid": "DecryptStream",
```

```
   "Effect": "Allow",
   "Action": [
    "kms:Decrypt"
   ],
   "Resource": "arn:aws:kms:us-
east-1:0123456789:key/1234abcd-12ab-34cd-56ef-1234567890ab"
  },
        {
   "Sid": "ListStream",
   "Effect": "Allow",
   "Action": "kinesis:ListStreams",
   "Resource": "*"
  }
 ]
}
```

2. Consulte la VPC y verifique que el clúster de Amazon Redshift o Amazon Redshift sin servidor dispone de una ruta para llegar a los puntos de conexión de Kinesis Data Streams a través de Internet mediante una puerta de enlace NAT o una puerta de enlace de Internet. Si desea que el tráfico entre Redshift y Kinesis Data Streams permanezca dentro de la red de AWS, considere la posibilidad de utilizar un punto de conexión de VPC de interfaz de Kinesis. Para obtener más información, consulte Using Amazon Kinesis Data Streams Kinesis Data Streams with Interface VPC Endpoints (Uso de Kinesis de Amazon Kinesis Data Streams con puntos de conexión de VPC de interfaz).
3. En Amazon Redshift, cree un esquema externo para asignar los datos de Kinesis a un esquema.

```
CREATE EXTERNAL SCHEMA kds
FROM KINESIS
IAM_ROLE { default | 'iam-role-arn' };
```

La ingesta de streaming para el flujo de datos de Kinesis no requiere ningún tipo de autenticación. Utiliza el rol de IAM definido en la instrucción `CREATE EXTERNAL SCHEMA` para realizar solicitudes de flujo de datos de Kinesis.

Opcional: Utilice la palabra clave REGION para especificar la región en la que reside la transmisión de Amazon Kinesis Data Streams o Amazon MSK.

```
CREATE EXTERNAL SCHEMA kds
FROM KINESIS
REGION 'us-west-2'
```

```
IAM_ROLE { default | 'iam-role-arn' };
```

En este ejemplo, la región especifica la ubicación del flujo de origen. El IAM_ROLE es un ejemplo.

4. Cree una vista materializada para consumir los datos del flujo. Con una instrucción como la siguiente, si un registro no se puede analizar, se produce un error. Use un comando como este si no quiere que se omitan los registros de errores.

```
CREATE MATERIALIZED VIEW my_view AUTO REFRESH YES AS
SELECT *
FROM kds.my_stream_name;
```

Los nombres de flujo de Kinesis distinguen entre mayúsculas y minúsculas y pueden contener letras tanto mayúsculas como minúsculas. Para la ingesta de flujos con nombres en mayúsculas, puede establecer la configuración `enable_case_sensitive_identifier` a `true` en el nivel de la base de datos. Para obtener más información, consulte Nombres e identificadores e enable_case_sensitive_identifier.

Para activar la actualización automática, utilice `AUTO REFRESH YES`. El comportamiento predeterminado es la actualización manual. Tenga en cuenta que cuando usa CAN_JSON_PARSE, es posible que se omitan los registros que no se pueden analizar.

Las columnas de metadatos incluyen lo siguiente:

| Columna de metadatos | Tipo de datos: | Descripción |
|---|---|---|
| approximate_arrival_timestamp | TIMESTAMP sin zona horaria | La hora aproximada en que se insertó el registro en el flujo de Kinesis |
| partition_key | varchar (256) | La clave que utiliza Kinesis para asignar el registro a una partición |
| shard_id | char(20) | El identificador único de la partición en el flujo del que se ha recuperado el registro |

| Columna de metadatos | Tipo de datos: | Descripción |
|---|---|---|
| sequence_number | varchar(128) | El identificador único del registro de la partición de Kinesis |
| refresh_time | TIMESTAMP sin zona horaria | La hora de inicio de la actualización |
| kinesis_data | varbyte | El registro del flujo de Kinesis |

Si tiene lógica empresarial en la definición de su vista materializada, es importante tener en cuenta que los errores de lógica empresarial pueden hacer que la ingesta de streaming se bloquee en algunos casos. Esto podría llevarlo a tener que eliminar y volver a crear la vista materializada. Para evitarlo, le recomendamos que mantenga su lógica lo más simple posible y realice la mayoría de las comprobaciones de lógica empresarial en los datos después de la ingesta.

5. Actualice la vista, lo que invoca a Redshift para que lea del flujo y cargue los datos en la vista materializada.

```
REFRESH MATERIALIZED VIEW my_view;
```

6. Consulte los datos en la vista materializada.

```
select * from my_view;
```

## Introducción a la ingesta de transmisiones desde orígenes de Apache Kafka

En este tema se describe cómo consumir datos de transmisiones de Amazon MSK, Apache Kafka o Confluent Cloud mediante una vista materializada.

El objetivo de la ingesta de streaming de Amazon Redshift es simplificar el proceso de ingesta directa de datos de flujo de un servicio de streaming en Amazon Redshift o Amazon Redshift sin servidor. Esto funciona con Amazon MSK aprovisionado y Amazon MSK sin servidor, con Apache Kafka de código abierto y con Confluent Cloud. La ingesta de transmisiones de Amazon Redshift elimina

la necesidad de preparar un tema de Apache Kafka en Amazon S3 antes de ingerir los datos de transmisiones en Redshift.

Desde un punto de vista técnico, la ingesta de transmisiones proporciona una ingesta de baja latencia y alta velocidad de datos de transmisiones o temáticos en una vista materializada de Amazon Redshift. Tras la configuración, si utiliza la actualización de vistas materializadas, podrá recibir grandes volúmenes de datos.

Debe disponer de un origen de Apache Kafka antes de configurar la ingesta de transmisiones de Amazon Redshift. Si no tiene un origen, cree uno mediante las siguientes instrucciones:

- Amazon MSK: Getting Started Using Amazon MSK
- Apache Kafka: Apache Kafka Quickstart
- Confluent Cloud: Quick Start for Confluent Cloud

## Configuración de la ingesta de transmisiones desde Kafka

Utilice los siguientes procedimientos para configurar la ingesta de transmisiones a Amazon Redshift desde Amazon MSK u orígenes de Apache Kafka que no estén administrados por AWS (Apache Kafka y Confluent Cloud).

Temas

- Configuración de la autenticación
- Configuración de la VPC
- Creación de una vista materializada

### Configuración de la autenticación

En esta sección se describe la configuración de la autenticación para permitir que la aplicación de Amazon Redshift acceda a un origen de Amazon MSK.

Tras crear el rol de la aplicación, asocie una de las siguientes políticas para permitir el acceso a su clúster de Amazon MSK, Apache Kafka o Confluent Cloud. Para la autenticación de mTLS, puede almacenar los certificados que utiliza Amazon Redshift en ACM o Secrets Manager, por lo que debe elegir la política correspondiente al lugar donde se almacena el certificado.

El control de acceso de IAM para Amazon Redshift está disponible para clústeres de MSK con la versión 2.7.1 de Kafka o superior.

AUTHENTICATION IAM (solo Amazon MSK):

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "MSKIAMpolicy",
            "Effect": "Allow",
            "Action": [
                "kafka-cluster:ReadData",
                "kafka-cluster:DescribeTopic",
                "kafka-cluster:Connect"
            ],
            "Resource": [
                "arn:aws:kafka:*:0123456789:cluster/MyTestCluster/*",
                "arn:aws:kafka:*:0123456789:topic/MyTestCluster/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "kafka-cluster:AlterGroup",
                "kafka-cluster:DescribeGroup"
            ],
            "Resource": [
                "arn:aws:kafka:*:0123456789:group/MyTestCluster/*"
            ]
        }
    ]
}
```

AUTHENTICATION MTLS: se utiliza un certificado almacenado en AWS Certificate Manager

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "MSKmTLSACMpolicy",
            "Effect": "Allow",
            "Action": [
                "acm:ExportCertificate"
            ],
            "Resource": [
```

```
                "arn:aws:acm:us-east-1:444455556666:certificate/certificate_ID"
            ]
        }
    ]
}
```

AUTHENTICATION MTLS: se utiliza un certificado almacenado en AWS Secrets Manager

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "MSKmTLSSecretsManagerpolicy",
            "Effect": "Allow",
            "Action": [
                "secretsmanager:GetSecretValue"
            ],
            "Resource": [
                "arn:aws:secretsmanager:us-east-1:444455556666:secret:secret_ID"
            ]
        }
    ]
}
```

Amazon MSK

Si utiliza AUTHENTICATION NONE para conectarse a un origen de Amazon MSK, no será necesario ningún rol de IAM. Sin embargo, si utiliza AUTHENTICATION IAM o MTLS para autenticarse con el clúster de Amazon MSK, el clúster de Amazon Redshift o el espacio de nombres de Amazon Redshift sin servidor deben tener un rol de IAM asociado con los permisos adecuados. Cree un rol de IAM con una política de confianza que permita que el clúster de Amazon Redshift o el espacio de nombres de Amazon Redshift sin servidor asuman el rol. Después de crear el rol, agregue uno de los siguientes permisos para admitir IAM o MTLS. Para la autenticación mTLS, los certificados que utiliza Amazon Redshift se pueden almacenar en AWS Certificate Manager o AWS Secrets Manager, por lo que debe elegir la política que coincida con el lugar donde se almacena el certificado. Asocie el rol al clúster de Amazon Redshift aprovisionado o un espacio de nombres de Redshift sin servidor. Para obtener información sobre cómo configurar la política de confianza del rol de IAM, consulte Autorización a Amazon Redshift para acceder a otros servicios de AWS en su nombre.

En la siguiente tabla se muestran las opciones de configuración complementarias a establecer para la ingesta de streaming desde Amazon MSK:

| Configuración de Amazon Redshift | Configuración de Amazon MSK | Puerto que se abrirá entre Redshift y Amazon MSK |
|---|---|---|
| AUTHENTICATION NONE | Transporte TLS desactivado | 9092 |
| AUTHENTICATION NONE | Transporte TLS habilitado | 9094 |
| AUTHENTICATION IAM | IAM | 9098/9198 |
| AUTHENTICATION MTLS | Transporte TLS habilitado | 9094 |

La autenticación de Amazon Redshift se establece en la instrucción CREATE EXTERNAL SCHEMA.

> **Note**
>
> En caso de que el clúster de Amazon MSK tenga activada la autenticación de seguridad de la capa de transporte mutua (mTLS), al configurar AUTHENTICATION NONE se indica a Amazon Redshift que utilice el puerto 9094 para el acceso sin autenticación. No obstante, se producirá un error porque el puerto lo está utilizando la autenticación mTLS. Por esto, le recomendamos que cambie a AUTHENTICATION mtls cuando use mTLS.

Apache Kafka or Confluent Cloud

Para Apache Kafka y Confluent Cloud, Amazon Redshift admite los siguientes protocolos de conexión:

- Puede utilizar mTLS o texto sin formato con transporte TLS para la autenticación cuando se conecte a Apache Kafka.
- Solo puede usar mTLS para la autenticación cuando se conecta a Confluent Cloud.

Amazon Redshift admite los siguientes protocolos de cifrado para conectarse a Apache Kafka o Confluent Cloud:

Métodos de autenticación compatibles con Apache Kafka y Confluent Cloud

| Amazon Redshift | Protocolo de seguridad de Kafka | Compatibilidad con Apache Kafka | Compatibilidad con Confluent Cloud |
|---|---|---|---|
| AUTHENTICATION NONE | PLAINTEXT | No | No |
| AUTHENTICATION NONE | SSL | Sí | No |
| AUTHENTICATION IAM | SASL_SSL | No | No |
| AUTHENTICATION MTLS | SSL | Sí (con certificado) | Sí (con certificado) |

Tenga en cuenta que Amazon Redshift no admite SASL/SCRAM ni SASL/PLAINTEXT.

Configuración de la VPC

Una vez que haya creado los recursos de autenticación, compruebe la VPC y verifique que el clúster de Amazon Redshift o el grupo de trabajo de Amazon Redshift sin servidor dispone de una ruta para llegar al origen de Apache Kafka.

> **Note**
>
> Para Amazon MSK, las reglas de grupo de seguridad de entrada del clúster de Amazon MSK deberían permitir el grupo de seguridad del clúster de Amazon Redshift o del grupo de trabajo de Redshift sin servidor. Los puertos que especifique dependerán de los métodos de autenticación configurados en el clúster de Amazon MSK. Para obtener más información, consulte Información de puertos y Acceso desde AWS pero fuera de la VPC.

A continuación, habilite el enrutamiento de VPC mejorado en el clúster de Amazon Redshift o grupo de trabajo de Amazon Redshift sin servidor. Para obtener más información, consulte Habilitación del enrutamiento de VPC mejorado.

### Creación de una vista materializada

En esta sección, configurará la vista materializada que Amazon Redshift utiliza para acceder a los datos de transmisiones de Apache Kafka.

Suponiendo que disponga de un clúster de Apache Kafka, el primer paso es definir un esquema en Redshift con `CREATE EXTERNAL SCHEMA` y hacer referencia al clúster como el origen de datos. Después de eso, para acceder a los datos del tema, defina `STREAM` en una vista materializada. Puede almacenar registros del tema con el tipo de datos VARBYTE de Amazon Redshift predeterminado o definir un esquema que convierte los datos en un formato `SUPER` semiestructurado. Cuando consulta la vista materializada, los registros devueltos son una vista de un punto temporal del tema.

1. En Amazon Redshift, cree un esquema externo para asignarlo al clúster de Apache Kafka. La sintaxis es la siguiente:

   ```
   CREATE EXTERNAL SCHEMA MySchema
   FROM KAFKA
   [ IAM_ROLE [ default | 'iam-role-arn' ] ]
   AUTHENTICATION [ none | iam | mtls ]
   {AUTHENTICATION_ARN 'acm-certificate-arn' |  SECRET_ARN 'asm-secret-arn'};
   ```

   En la cláusula `FROM`, `KAFKA` indica que el esquema asigna los datos desde un origen de Apache Kafka.

   `AUTHENTICATION` indica el tipo de autenticación para la ingesta de transmisiones. Hay tres tipos disponibles:

   - ninguno: especifica que no se requiere autenticación. Esto corresponde al acceso no autenticado en MSK. Esto corresponde a la autenticación SSL en Apache Kafka. Este método de autenticación no es compatible con Confluent Cloud.
   - iam: especifica la autenticación de IAM. Solo puede utilizar la autenticación de IAM con Amazon MSK. Al elegir esta opción, asegúrese de que el rol de IAM tenga permisos para la autenticación de IAM. Para obtener más información sobre la configuración de las políticas de IAM requeridas, consulte Configuración de la ingesta de transmisiones desde Kafka.
   - mtls: especifica que la seguridad de la capa de transporte mutua proporciona una comunicación segura al facilitar la autenticación entre un cliente y un servidor. En este caso, el cliente es Redshift y el servidor es Apache Kafka. Para obtener más información acerca de

cómo configurar la ingesta de streaming con mTLS, consulte Autenticación con mTLS para la ingesta de transmisiones de Redshift desde orígenes de Apache Kafka.

Tenga en cuenta que la autenticación de Amazon MSK con un nombre de usuario y la contraseña no es compatible con la ingesta de streaming.

El parámetro AUTHENTICATION_ARN especifica el ARN del certificado de seguridad de la capa de transporte mutua (mTLS) de ACM que se utiliza para establecer una conexión cifrada.

El parámetro SECRET_ARN especifica el arn del secreto de AWS Secrets Manager que contiene el certificado que utilizará Amazon Redshift para mTLS.

En los siguientes ejemplos se muestra cómo establecer el URI del agente y el clúster de Amazon MSK al crear el esquema externo:

Uso de la autenticación de IAM:

```
CREATE EXTERNAL SCHEMA my_schema
FROM KAFKA
IAM_ROLE 'arn:aws:iam::012345678901:role/my_role'
AUTHENTICATION IAM
URI 'b-1.myTestCluster.123z8u.c2.kafka.us-
west-1.amazonaws.com:9098,b-2.myTestCluster.123z8u.c2.kafka.us-
west-1.amazonaws.com:9098'
```

Uso sin autenticación:

```
CREATE EXTERNAL SCHEMA my_schema
FROM KAFKA
AUTHENTICATION none
URI 'b-1.myTestCluster.123z8u.c2.kafka.us-
west-1.amazonaws.com:9092,b-2.myTestCluster.123z8u.c2.kafka.us-
west-1.amazonaws.com:9092'
```

Uso de mTLS:

```
CREATE EXTERNAL SCHEMA my_schema
FROM KAFKA
IAM_ROLE 'arn:aws:iam::012345678901:role/my_role'
AUTHENTICATION MTLS
```

```
URI 'b-1.myTestCluster.123z8u.c2.kafka.us-west-1.amazonaws.com:9094,b-
 2.myTestCluster.123z8u.c2.kafka.us-west-1.amazonaws.com:9094'
{AUTHENTICATION_ARN 'acm-certificate-arn' |  SECRET_ARN 'asm-secret-arn'}
```

Para obtener más información sobre la creación de un esquema externo, consulte CREATE EXTERNAL SCHEMA.

2. Cree una vista materializada para consumir los datos del tema. Utilice un comando SQL como el del ejemplo siguiente.

```
CREATE MATERIALIZED VIEW MyView AUTO REFRESH YES AS
SELECT *
FROM MySchema."mytopic";
```

En los nombres de tema de Kafka se distingue entre mayúsculas y minúsculas, y pueden estar en mayúsculas y minúsculas. Para la ingesta de temas con nombres en mayúsculas, puede establecer la configuración `enable_case_sensitive_identifier` en `true` en el nivel de sesión o de base de datos. Para obtener más información, consulte Nombres e identificadores e enable_case_sensitive_identifier.

Para activar la actualización automática, utilice `AUTO REFRESH YES`. El comportamiento predeterminado es la actualización manual.

3. Las columnas de metadatos incluyen lo siguiente:

| Columna de metadatos | Tipo de datos: | Descripción |
|---|---|---|
| kafka_partition | bigint | Id. de partición del registro del tema de Kafka |
| kafka_offset | bigint | Desplazamiento del registro en el tema de Kafka para una partición determinada |
| kafka_timestamp_type | char(1) | Tipo de la marca temporal utilizada en el registro de Kafka: |

| Columna de metadatos | Tipo de datos: | Descripción |
|---|---|---|
| | | • C: hora de creación del registro (CREATE_TIME) en el cliente<br>• L: hora de adición del registro (LOG_APPEND_TIME) en el servidor Kafka<br>• U: hora de creación del registro no está disponible (NO_TIMESTAMP_TYPE) |
| kafka_timestamp | TIMESTAMP sin zona horaria | Valor de la marca temporal del registro |
| kafka_key | varbyte | Clave del registro de Kafka |
| kafka_value | varbyte | Registro recibido de Kafka |
| kafka_headers | super | Encabezado del registro recibido de Kafka |
| refresh_time | TIMESTAMP sin zona horaria | La hora de inicio de la actualización |

Es importante tener en cuenta, si tiene lógica empresarial en su definición de vista materializada, que los resultados en los errores de lógica empresarial pueden hacer que la ingesta de streaming produzca errores en algunos casos. Esto podría llevarlo a tener que eliminar y volver a crear la vista materializada. Para evitarlo, le recomendamos que mantenga su lógica empresarial simple y que aplique lógica adicional a los datos después de la ingesta.

4. Actualice la vista, lo que invoca a Amazon Redshift para que lea del tema y cargue los datos en la vista materializada.

```
REFRESH MATERIALIZED VIEW MyView;
```

5. Consulte los datos en la vista materializada.

```
select * from MyView;
```

La vista materializada se actualiza directamente desde el tema cuando se ejecuta REFRESH. Se crea una vista materializada que se asigna al origen de datos del tema de Kafka. Puede realizar filtrados y agregaciones en los datos como parte de la definición de la vista materializada. La vista materializada de ingesta de streaming (vista materializada base) solo puede hacer referencia a un tema de Kafka, pero se pueden crear vistas materializadas adicionales que se unan con la vista materializada base y con otras vistas materializadas o tablas.

Para obtener más información sobre las limitaciones para la ingesta de streaming, consulte Comportamiento y tipos de datos de ingesta de streaming.

## Autenticación con mTLS para la ingesta de transmisiones de Redshift desde orígenes de Apache Kafka

La seguridad mutua de la capa de transporte (mTLS) proporciona los medios para que un servidor autentique a un cliente al que envía información y para que el cliente autentique el servidor. El beneficio de usar mTLS es que proporciona una autenticación fiable para una variedad de casos de uso en varias aplicaciones del sector vertical. Estos incluyen casos de uso en los sectores financiero, minorista, gubernamental y sanitario. En el caso de la ingesta de transmisiones en Redshift, la autenticación se produce entre un servidor, que puede ser Amazon MSK, Apache Kafka o Confluent Cloud, y un clúster aprovisionado de Amazon Redshift o un grupo de trabajo de Amazon Redshift sin servidor.

En este tema se proporcionan procedimientos y ejemplos de comandos de SQL que muestran formas de crear un esquema externo que utiliza mTLS para autenticarse entre el cliente de Redshift y cualquier servidor de Apache Kafka. Los pasos de este tema complementan el conjunto completo de pasos para configurar la ingesta de transmisiones desde orígenes de Apache Kafka. Para obtener más información, consulte Introducción a la ingesta de transmisiones desde orígenes de Apache Kafka.

Requisitos previos para usar mTLS para la ingesta de streaming

En esta sección se proporcionan los pasos previos para usar mTLS para la ingesta de transmisiones con AWS Certificate Manager o AWS Secrets Manager.

Como paso preliminar, debe tener o crear una autoridad de certificación privada (PCA), que puede utilizar para emitir certificados que, entre otras funciones, permiten una comunicación segura a través

de canales de comunicación seguros. AWS Private Certificate Authority (Entidad de certificación privada) es un servicio disponible que realiza esta función. Para obtener más información, consulte Creación de una entidad de certificación privada en la Guía del usuario de AWS Private Certificate Authority. Tras crear la CA privada, exporte el certificado de CA raíz y guárdelo en un archivo con la extensión .pem.

Para crear un clúster que utilice el certificado de CA, haga lo siguiente:

Amazon MSK

1. Cree un clúster de Amazon MSK que admita la autenticación de clientes mtls. Para obtener más información sobre la configuración de un clúster de Amazon MSK, consulte Autenticación TLS mutua de clientes para Amazon MSK en la Guía para desarrolladores de Amazon Managed Streaming para Apache Kafka.
2. Edite la configuración de seguridad del clúster de Amazon MSK, active la autenticación de cliente de TLS mediante AWS Certificate Manager (ACM) y seleccione la AWS Private CA (PCA) que creó anteriormente. Para obtener más información, consulte Actualización de la configuración de seguridad de un clúster en la Guía para desarrolladores de Amazon Managed Streaming para Apache Kafka.

Confluent Cloud

1. Cree un clúster de Confluent Cloud dedicado, preferiblemente en la misma Región de AWS que el clúster de Amazon Redshift. Para obtener información sobre cómo crear un clúster de Confluent Cloud, consulte Create a Kafka cluster in Confluent Cloud.
2. Cargue el archivo pem del certificado de CA raíz de AWS Private CA exportado que creó anteriormente. Para obtener más información, consulte Manage certificate authority for mTLS authentication for Confluent Cloud. Confluent Cloud usa este certificado para verificar el certificado de cliente de Amazon Redshift.

Uso de mTLS para la ingesta de datos con AWS Certificate Manager

El siguiente procedimiento muestra cómo configurar mTLS para la ingesta de transmisiones de Redshift utilizando AWS Certificate Manager (ACM) para el almacenamiento y la administración de certificados:

1. Solicite un certificado privado a través de ACM. Cuando lo haga, seleccione la PCA que creó en la sección de requisitos previos como autoridad de certificación. ACM almacena el

certificado firmado y la clave privada adjunta para una comunicación segura. Para obtener más información sobre la administración de certificados con ACM, consulte Emisión y administración de certificados en la Guía del usuario de AWS Certificate Manager.

2. Para el rol de IAM que utilice para administrar el clúster de Redshift o el grupo de trabajo de Amazon Redshift sin servidor, adjunte el permiso para exportar el certificado, que es acm:ExportCertificate. Para obtener más información acerca de cómo configurar los recursos de IAM necesarios para la ingesta de transmisiones, consulte Configuración de la ingesta de transmisiones desde Kafka. Especifique el mismo rol de IAM en el siguiente paso para crear el esquema externo.

   > **Note**
   >
   > Solicitudes para que AWS Certificate Manager requiera una puerta de enlace de Internet (IGW) o una puerta de enlace NAT (NGW) en su VPC. Si su VPC no tiene una IGW ni una NGW, haga lo siguiente:
   >
   > - Utilice Secrets Manager en lugar de ACM para almacenar sus certificados.
   > - Adjunte un punto de conexión de VPC de Secrets Manager a su VPC.
   >
   > Para obtener información sobre el uso de Secrets Manager con mTLS para la ingesta de streaming, consulte Uso de mTLS para la ingesta de datos con AWS Secrets Manager.

3. Obtenga el URI del agente de arranque para el clúster de Amazon MSK, Apache Kafka o Confluent Cloud. Para obtener información sobre cómo obtener el URI del agente de arranque para Amazon MSK, consulte Obtención de agentes de arranque para un clúster de Amazon MSK en la Guía para desarrolladores de Amazon Managed Streaming para Apache Kafka.

4. Ejecute un comando de SQL como en el siguiente ejemplo para crear un esquema externo que asigne el clúster a un esquema externo de Redshift utilizando `mtls`.

   Amazon MSK

```
CREATE EXTERNAL SCHEMA my_schema
FROM KAFKA
IAM_ROLE 'arn:aws:iam::012345678901:role/my_role'
AUTHENTICATION mtls
URI 'b-1.myTestCluster.123z8u.c2.kafka.us-
west-1.amazonaws.com:9094,b-2.myTestCluster.123z8u.c2.kafka.us-
west-1.amazonaws.com:9094'
```

```
AUTHENTICATION_ARN 'arn:aws:acm:Region:444455556666:certificate/certificate_ID';
```

Apache Kafka or Confluent Cloud

```
CREATE EXTERNAL SCHEMA my_schema
FROM KAFKA
IAM_ROLE 'arn:aws:iam::012345678901:role/my_role'
AUTHENTICATION mtls
URI 'lkc-2v531.domz6wj0p.us-west-1.aws.confluent.cloud:9092'
AUTHENTICATION_ARN 'arn:aws:acm:region:444455556666:certificate/certificate_ID';
```

Parámetros importantes:

- IAM_ROLE: el rol de IAM asociado al clúster, para la ingesta de streaming.
- URI: URI del agente de arranque para el clúster. Tenga en cuenta que, para Amazon MSK, se especifica el puerto 9094 para comunicarse con los agentes de cifrado TLS.
- AUTHENTICATION_ARN: el ARN del certificado de ACM. El ARN está disponible en la consola de ACM al elegir el certificado emitido.

Tras realizar estos pasos de configuración, puede crear una vista materializada de Redshift que haga referencia al esquema definido en el ejemplo y, a continuación, utilizar REFRESH MATERIALIZED VIEW para transmitir los datos. Para obtener más información, consulte Introducción a la ingesta de transmisiones desde orígenes de Apache Kafka.

Uso de mTLS para la ingesta de datos con AWS Secrets Manager

Puede configurar mTLS para la ingesta de transmisiones de Redshift utilizando AWS Secrets Manager para la administración de certificados si no desea hacer referencia al certificado en AWS Certificate Manager. En los siguientes pasos se describe cómo configurar los mTLS mediante Secrets Manager.

1. Cree una solicitud de firma de certificado y una clave privada con la herramienta que prefiera. A continuación, puede utilizar la solicitud de firma para generar un certificado firmado con la misma entidad de certificación privada de AWS (PCA) que utilizó para generar el certificado para el clúster. Para obtener más información sobre la emisión de un certificado, consulte IssueCertificate en la Referencia de la API de AWS Private Certificate Authority.

2. Extraiga el certificado mediante AWS Private Certificate Authority. Para obtener más información, consulte Retrieve a private certificate en la Guía del usuario de AWS Private Certificate Authority.

3. Guarde el certificado y la clave privada generados en el paso anterior enAWS Secrets Manager. Elija `Other type of secret` y utilice el formato de texto plano. Los pares clave-valor deben tener el formato `{"certificate":"<cert value>","privateKey":"<pkey value>"}`, que se muestra en el siguiente ejemplo. Para obtener más información acerca de la creación y administración de secretos en AWS Secrets Manager, consulte Creación y administración de secretos con AWS Secrets Manager en la Guía del usuario de AWS Secrets Manager.

```
{"certificate":"-----BEGIN CERTIFICATE-----
klhdslkfjahksgdfkgioeuyihbflahabhbdslv6akybeoiwv1hoaiusdhbahsbdi
H4hAX8/eE96qCcjkpfT84EdvHzp6fC+/WwM0oXlwUEWlvfMCXNaG5D8SqRq3qA==
-----END CERTIFICATE-----
-----BEGIN CERTIFICATE-----
klhdslkfjahksgdfkgioeuyihbflahabhbdslv6akybeoiwv1hoaiusdhbahsbdi
wJalrXUtnFEMI/K7MDENG/bPxRfiCYEXAMPLEKEY
-----END CERTIFICATE-----",
"privateKey":"-----BEGIN PRIVATE KEY-----
klhdslkfjahksgdfkgioeuyihbflahabhbdslv6akybeoiwv1hoaiusdhbahsbdi
7OD4m1dBEs3Fj++hDMH9rYRp99RqtCOf0EWOUe139KOilOsW+cyhAoc9Ci2+Jo/k
17u2N1iGILMQEZuCRtnJOkFYkw==
-----END PRIVATE KEY-----"}
```

4. Asocie la política de permisos para recuperar el secreto para el rol de IAM que utiliza para administrar el clúster de Amazon Redshift o el grupo de trabajo de Amazon Redshift sin servidor. Este permiso es `secretsmanager:GetSecretValue`. Para obtener más información, consulte Configuración de la autenticación. Para obtener más información sobre la administración de las políticas de IAM, consulte Edición de políticas de IAM. Especifique el mismo rol de IAM en el siguiente paso para crear el esquema externo.

5. En Redshift, ejecute el comando SQL para crear el esquema externo. Se utiliza el tipo de AUTENTICACIÓN `mtls`. También se debe especificar el URI del clúster y el ARN del secreto en AWS Secrets Manager.

```
CREATE EXTERNAL SCHEMA my_schema
FROM KAFKA
IAM_ROLE 'arn:aws:iam::012345678901:role/my_role'
AUTHENTICATION mtls
```

```
URI 'b-1.myTestCluster.123z8u.c2.kafka.us-
west-1.amazonaws.com:9094,b-2.myTestCluster.123z8u.c2.kafka.us-
west-1.amazonaws.com:9094'
SECRET_ARN 'arn:aws:secretsmanager:us-east-1:012345678910:secret:myMTLSSecret';
```

Parámetros importantes:

- IAM_ROLE: el rol de IAM asociado al clúster, para la ingesta de streaming.
- URI: URI del agente de arranque para el clúster. Tenga en cuenta que, para Amazon MSK, se especifica el puerto 9094 para comunicarse con los agentes de cifrado TLS.
- SECRET_ARN: el ARN del secreto de Secrets Manager, que contiene el certificado que se utilizará para mTLS.

Habilitación de la autenticación mTLS para un esquema externo existente

Si tiene un esquema externo existente que utiliza para la ingesta de streaming y desea implementar un TLS mutuo para la autenticación, puede ejecutar un comando como el siguiente, que especifica la autenticación mTLS y el ARN del certificado ACM en ACM.

```
ALTER EXTERNAL SCHEMA schema_name
AUTHENTICATION mtls
AUTHENTICATION_ARN 'arn:aws:acm:Region:444455556666:certificate/certificate_ID';
```

O bien, puede especificar la autenticación mTLS, con referencia al ARN del secreto en AWS Secrets Manager.

```
ALTER EXTERNAL SCHEMA schema_name
AUTHENTICATION mtls
SECRET_ARN 'arn:aws:secretsmanager:us-east-1:012345678910:secret:myMTLSSecret';
```

Para obtener información sobre el comando ALTER EXTERNAL SCHEMA, consulte ALTER EXTERNAL SCHEMA.

## Ingesta de datos de streaming mediante Kinesis

Este procedimiento muestra cómo ingerir datos de un flujo de Kinesis denominado ev_station_data, que contiene datos de consumo de diferentes estaciones de carga de vehículos eléctricos, en formato JSON. El esquema está bien definido. El ejemplo muestra cómo almacenar los datos en

formato JSON sin procesar, y también cómo convertir los datos JSON a tipos de datos de Amazon Redshift a medida que se ingieren.

Configuración del productor

1. Utilizando Amazon Kinesis Data Streams, siga los pasos para crear un flujo denominado `ev_station_data`. Elija On-demand (Bajo demanda) para Capacity mode (Modo de capacidad). Para obtener más información, consulte Creación de un flujo a través de la consola de administración de AWS.
2. El Generador de datos de Amazon Kinesis puede ayudarle a generar datos de prueba para utilizarlos con el flujo. Siga los pasos detallados en la herramienta para comenzar, y utilice la siguiente plantilla de datos para generar los datos:

```
{

   "_id" : "{{random.uuid}}",
   "clusterID": "{{random.number(
        {   "min":1,
            "max":50
        }
    )}}",
    "connectionTime": "{{date.now("YYYY-MM-DD HH:mm:ss")}}",
    "kWhDelivered": "{{commerce.price}}",
    "stationID": "{{random.number(
        {   "min":1,
            "max":467
        }
    )}}",
      "spaceID": "{{random.word}}-{{random.number(
        {   "min":1,
            "max":20
        }
    )}}",

   "timezone": "America/Los_Angeles",
   "userID": "{{random.number(
        {   "min":1000,
            "max":500000
        }
    )}}"
}
```

Cada objeto JSON de los datos del flujo tiene las siguientes propiedades:

```
{
    "_id": "12084f2f-fc41-41fb-a218-8cc1ac6146eb",
    "clusterID": "49",
    "connectionTime": "2022-01-31 13:17:15",
    "kWhDelivered": "74.00",
    "stationID": "421",
    "spaceID": "technologies-2",
    "timezone": "America/Los_Angeles",
    "userID": "482329"
}
```

### Configuración de Amazon Redshift

Estos pasos muestran cómo configurar la vista materializada para ingerir datos.

1. Cree un esquema externo para asignar los datos de Kinesis a un objeto de Redshift.

   ```
   CREATE EXTERNAL SCHEMA evdata FROM KINESIS
   IAM_ROLE 'arn:aws:iam::0123456789:role/redshift-streaming-role';
   ```

   Para obtener más información sobre cómo configurar el rol de IAM, consulte Introducción a la ingesta de streaming de Amazon Kinesis Data Streams.

2. Cree una vista materializada para consumir los datos del flujo. En el siguiente ejemplo se muestra cómo definir una vista materializada para ingerir los datos en formato JSON de un flujo de Kinesis.

   En primer lugar, almacene los registros del flujo en formato SUPER semiestructurado. En este ejemplo, el origen JSON se almacena en Redshift sin realizar conversión a tipos de Redshift.

   ```
   CREATE MATERIALIZED VIEW ev_station_data AS
       SELECT approximate_arrival_timestamp,
       partition_key,
       shard_id,
       sequence_number,
       case when can_json_parse(kinesis_data) then json_parse(kinesis_data) else null
    end as payload,
   ```

```
   case when not can_json_parse(kinesis_data) then kinesis_data else null end as
 failed_payload
   FROM evdata."ev_station_data" ;
```

Consulta del flujo

1. Active atributos SUPER que distingan entre mayúsculas y minúsculas mediante el siguiente comando. De forma predeterminada, Amazon Redshift no distingue entre mayúsculas y minúsculas, por lo que para acceder a los atributos SUPER que distingan entre mayúsculas y minúsculas, debe habilitar esta funcionalidad.

   ```
   SET enable_case_sensitive_super_attribute to TRUE;
   ```

2. Actualice la vista materializada con el siguiente comando para extraer datos del flujo.

   ```
   REFRESH MATERIALIZED VIEW ev_station_data;
   ```

3. Consulte la vista materializada actualizada para obtener estadísticas de uso.

   ```
   SELECT e.payload.connectionTime::date as connectiontime
   ,SUM(e.payload.kWhDelivered::decimal(10,2)) AS Energy_Consumed
   ,count(distinct e.payload.userID) AS #Users
   from ev_station_data as e
   group by connectiontime
   order by 1 desc;
   ```

4. Vea los resultados.

   ```
   connectiontime        energy_consumed    #users
   2022-02-08                 4139          10
   2022-02-09                 5571          10
   2022-02-10                 8697          20
   2022-02-11                 4408          10
   2022-02-12                 4257          10
   2022-02-23                 6861          10
   ```

# Vistas de AWS Glue Data Catalog

En este tema se describe cómo crear vistas en AWS Glue Data Catalog. Puede usar las vistas del catálogo de datos para acceder a los datos de diferentes orígenes de datos mediante el mismo esquema.

Al crear vistas en el catálogo de datos, puede crear un único objeto de metadatos y esquema de vista común para usarlo en motores como Amazon Athena y Amazon EMR Spark. De este modo, podrá utilizar las mismas vistas en todos sus lagos y almacenamientos de datos según sus casos prácticos. Las vistas del catálogo de datos son especiales porque se categorizan como vistas de definidor, donde los permisos de acceso los define el usuario que creó la vista y no el usuario que consulta la vista. A continuación se muestran algunos casos prácticos y ventajas de crear vistas en el catálogo de datos:

- Cree una vista que restrinja el acceso a los datos en función de los permisos que necesite el usuario. Por ejemplo, puede usar las vistas del catálogo de datos para evitar que el personal que no trabaje en el departamento de Recursos Humanos (RR.HH.) vea información de identificación personal (PII).
- Asegúrese de que los usuarios no puedan acceder a registros incompletos. Al aplicar determinados filtros a la vista del catálogo de datos, se asegura de que los registros de datos de una vista del catálogo de datos estén siempre completos.
- Las vistas del catálogo de datos incluyen la ventaja de seguridad, ya que garantizan que la definición de consulta utilizada para crear la vista deba completarse para poder crearla. Esta ventaja de seguridad significa que las vistas del catálogo de datos no son susceptibles a los comandos SQL de actores malintencionados.
- Las vistas del catálogo de datos ofrecen las mismas ventajas que las vistas normales, como permitir a los usuarios acceder a una vista sin poner a su disposición la tabla subyacente.

Para crear una vista en el catálogo de datos, debe tener una tabla externa de Spectrum, un objeto incluido en un recurso compartido de datos administrado por Lake Formation o una tabla de Apache Iceberg.

Las definiciones de las vistas del catálogo de datos se almacenan en el AWS Glue Data Catalog. Use AWS Lake Formation para conceder el acceso mediante concesiones de recursos, concesiones de columnas o controles de acceso basados en etiquetas. Para obtener más información sobre cómo

conceder y revocar el acceso a Lake Formation, consulte Concesión y revocación de permisos sobre los recursos del catálogo de datos.

## Requisitos previos

Antes de crear una vista en el catálogo de datos, asegúrese de que cumple los siguientes requisitos previos:

- Asegúrese de que su rol de IAM tenga la siguiente política de confianza.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
               "Service": [
                    "glue.amazonaws.com",
                    "lakeformation.amazonaws.com"
                ]
            },
            "Action": "sts:AssumeRole"
        }
    ]
}
```

- También necesita la siguiente política para transmitir roles.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "Stmt1",
            "Action": [
                "iam:PassRole"
            ],
            "Effect": "Allow",
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "iam:PassedToService": [
                        "glue.amazonaws.com",
```

```
                        "lakeformation.amazonaws.com"
                    ]
                }
            }
        }
    ]
}
```

- Por último, también necesita los siguientes permisos.
  - `Glue:GetDatabase`
  - `Glue:GetDatabases`
  - `Glue:CreateTable`
  - `Glue:GetTable`
  - `Glue:UpdateTable`
  - `Glue:DeleteTable`
  - `Glue:GetTables`
  - `Glue:SearchTables`
  - `Glue:BatchGetPartition`
  - `Glue:GetPartitions`
  - `Glue:GetPartition`
  - `Glue:GetTableVersion`
  - `Glue:GetTableVersions`

# Ejemplo completo

Para empezar, cree un esquema externo basado en la base de datos del catálogo de datos.

```
CREATE EXTERNAL SCHEMA IF NOT EXISTS external_schema FROM DATA CATALOG DATABASE
 'external_data_catalog_db'
IAM_ROLE 'arn:aws:iam::123456789012:role/sample-role';
```

Ahora puede crear una vista de catálogo de datos.

```
CREATE EXTERNAL PROTECTED VIEW external_schema.remote_view
AS SELECT * FROM external_schema.remote_table;
```

A continuación, puede empezar a realizar consultas a la vista.

```
SELECT * FROM external_schema.remote_view;
```

Para obtener más información sobre los comandos SQL relacionados con las vistas del catálogo de datos, consulte CREATE EXTERNAL VIEW, ALTER EXTERNAL VIEW y DROP EXTERNAL VIEW.

## Registro seguro

Redshift enmascara los metadatos guardados en los registros del sistema de Redshift cuando una consulta hace referencia a una vista Glue multidialectal. Multidialectal significa que la vista admite los dialectos SQL de distintos motores de consulta, como Redshift y Amazon EMR. Los datos de las tablas siguientes están enmascarados para todas las consultas con el mismo ID de consulta. En la siguiente tabla se enumeran las vistas del sistema y las columnas con el registro seguro aplicado.

| Tabla del sistema | Columnas confidenciales |
|---|---|
| SYS_EXTERNAL_QUERY_DETAIL | Columnas: source_type, total_partitions, qualified_partitions, scanned_files, returned_rows, returned_bytes, file_format, file_location, external_query_text y warning_message. Para obtener más información, consulte SYS_EXTERNAL_QUERY_DETAIL. |
| SYS_EXTERNAL_QUERY_ERROR | Columnas: file_location, rowid, column_name, original_value, modified_value, trigger, action, action_value y error_code. Para obtener más información, consulte SYS_EXTERNAL_QUERY_ERROR. |
| SYS_QUERY_DETAIL | Columnas: step_name, table_id, table_name, input_bytes, input_rows, output_bytes, output_rows, blocks_read, blocks_write, local_read_IO, remote_read_IO, spilled_block_local_disk y spilled_block_remote_disk. Para obtener más información, consulte SYS_QUERY_DETAIL. |
| SYS_QUERY_HISTORY | Columnas: returned_rows y returned_bytes. Para obtener más información, consulte SYS_QUERY_HISTORY. |
| STL_AGGR | Columnas: rows, bytes, tbl y type. Para obtener más información, consulte STL_AGGR. |

| Tabla del sistema | Columnas confidenciales |
|---|---|
| STL_BCAST | Columnas: rows, bytes y packets. Para obtener más información, consulte STL_BCAST. |
| STL_DDLTEXT | Columnas: text. Para obtener más información, consulte STL_DDLTEXT. |
| STL_DELETE | Columnas: rows y tbl. Para obtener más información, consulte STL_DELETE. |
| STL_DIST | Columnas: rows, bytes y packets. Para obtener más información, consulte STL_DIST. |
| STL_EXPLAIN | Columnas: plannode e info. Para obtener más información, consulte STL_EXPLAIN. |
| STL_HASH | Columnas: rows, bytes, tbl y est_rows. Para obtener más información, consulte STL_HASH. |
| STL_HASHJOIN | Columnas: rows, tbl, num_parts y join_type. Para obtener más información, consulte STL_HASHJOIN. |
| STL_INSERT | Columnas: rows y tbl. Para obtener más información, consulte STL_INSERT. |
| STL_LIMIT | Columnas: rows. Para obtener más información, consulte STL_LIMIT. |
| STL_MERGE | Columnas: rows. Para obtener más información, consulte STL_MERGE. |
| STL_MERGEJOIN | Columnas: rows y tbl. Para obtener más información, consulte STL_MERGEJOIN. |
| STL_NESTLOOP | Columnas: rows y tbl. Para obtener más información, consulte STL_NESTLOOP. |
| STL_PARSE | Columnas: rows. Para obtener más información, consulte STL_PARSE. |

| Tabla del sistema | Columnas confidenciales |
|---|---|
| STL_PLAN_INFO | Columnas: rows y bytes. Para obtener más información, consulte STL_PLAN_INFO. |
| STL_PROJECT | Columnas: rows y tbl. Para obtener más información, consulte STL_PROJECT. |
| STL_QUERY | Columnas: querytxt. Para obtener más información, consulte STL_QUERY. |
| STL_QUERY_METRICS | Columnas: max_rows, rows, max_blocks_read, blocks_read, max_blocks_to_disk, blocks_to_disk, max_query_scan_size y query_scan_size. Para obtener más información, consulte STL_QUERY_METRICS. |
| STL_QUERYTEXT | Columnas: text. Para obtener más información, consulte STL_QUERY TEXT. |
| STL_RETURN | Columnas: rows y bytes. Para obtener más información, consulte STL_RETURN. |
| STL_SAVE | Columnas: rows, bytes y tbl. Para obtener más información, consulte STL_SAVE. |
| STL_SCAN | Columnas: rows, bytes, fetches, type, tbl, rows_pre_filter, perm_tabl e_name y scanned_mega_value. Para obtener más información, consulte STL_SCAN. |
| STL_SORT | Columnas: rows, bytes y tbl. Para obtener más información, consulte STL_SORT. |
| STL_TR_CONFLICT | Columnas: table_id. Para obtener más información, consulte STL_TR_CONFLICT. |
| STL_UNDONE | Columnas: table_id. Para obtener más información, consulte STL_UNDONE. |

| Tabla del sistema | Columnas confidenciales |
|---|---|
| STL_UNIQUE | Columnas: rows, type y bytes. Para obtener más información, consulte STL_UNIQUE. |
| STL_UTILITYTEXT | Columnas: text. Para obtener más información, consulte STL_UTILITYTEXT. |
| STL_WINDOW | Columnas: rows. Para obtener más información, consulte STL_WINDOW. |
| STV_BLOCKLIST | Columnas: col, tbl, num_values, minvalue y maxvalue. Para obtener más información, consulte STV_BLOCKLIST. |
| STV_EXEC_STATE | Columnas: rows, bytes y label. Para obtener más información, consulte STV_EXEC_STATE. |
| STV_LOCKS | Columnas: table_id. Para obtener más información, consulte STV_LOCKS. |
| STV_QUERY_METRICS | Columnas: rows, max_rows, blocks_read, max_blocks_read, max_blocks_to_disk, blocks_to_disk, max_query_scan_size y query_scan_size. Para obtener más información, consulte STV_QUERY_METRICS. |
| STV_STARTUP_RECOVERY_STATE | Columnas: table_id y table_name. Para obtener más información, consulte STV_STARTUP_RECOVERY_STATE. |
| STV_TBL_PERM | Columnas: id, rows, sorted_rows, temp, block_count y query_scan_size. Para obtener más información, consulte STV_TBL_PERM. |
| STV_TBL_TRANS | Columnas: id, rows y size. Para obtener más información, consulte STV_TBL_TRANS. |
| SVCS_EXPLAIN | Columnas: plannode e info. Para obtener más información, consulte SVCS_EXPLAIN. |

| Tabla del sistema | Columnas confidenciales |
|---|---|
| SVCS_PLAN_INFO | Columnas: rows y bytes. Para obtener más información, consulte SVCS_PLAN_INFO. |
| SVCS_QUERY_SUMMARY | Columnas: step, rows, bytes, rate_row, rate_byte, label y rows_pre_filter. Para obtener más información, consulte SVCS_QUERY_SUMMARY. |
| SVCS_S3LIST | Columnas: bucket, prefix, max_file_size y avg_file_size. Para obtener más información, consulte SVCS_QUERY_SUMMARY. |
| SVCS_S3LOG | Columnas: message. Para obtener más información, consulte SVCS_QUERY_SUMMARY. |
| SVCS_S3PARTITION_SUMMARY | Columnas: total_partitions, qualified_partitions, min_assigned_partitions, max_assigned_partitions y avg_assigned_partitions. Para obtener más información, consulte SVCS_S3PARTITION_SUMMARY. |
| SVCS_S3QUERY_SUMMARY | Columnas: external_table_name, file_format, s3_scanned_rows, s3_scanned_bytes, s3query_returned_rows y s3query_returned_bytes. Para obtener más información, consulte SVCS_S3QUERY_SUMMARY. |
| SVL_QUERY_METRICS | Columnas: step_label, scan_row_count, join_row_count, nested_loop_join_row_count, return_row_count, spectrum_scan_row_count y spectrum_scan_size_mb. Para obtener más información, consulte SVL_QUERY_METRICS. |
| SVL_QUERY_METRICS_SUMMARY | Columnas: step_label, scan_row_count, join_row_count, nested_loop_join_row_count, return_row_count, spectrum_scan_row_count y spectrum_scan_size_mb. Para obtener más información, consulte SVL_QUERY_METRICS_SUMMARY. |
| SVL_QUERY_REPORT | Columnas: rows, bytes, label y rows_pre_filter. Para obtener más información, consulte SVL_QUERY_REPORT. |
| SVL_QUERY_SUMMARY | Columnas: rows, bytes y rows_pre_filter. Para obtener más información, consulte SVL_QUERY_SUMMARY. |

| Tabla del sistema | Columnas confidenciales |
|---|---|
| SVL_S3LIST | Columnas: bucket, prefix, max_file_size y avg_file_size. Para obtener más información, consulte SVL_S3LIST. |
| SVL_S3LOG | Columnas: message. Para obtener más información, consulte SVL_S3LOG. |
| SVL_S3PARTITION | Columnas: rows, bytes, label y rows_pre_filter. Para obtener más información, consulte SVL_S3PARTITION. |
| SVL_S3PARTITION_SUMMARY | Columnas: total_partitions, qualified_partitions, min_assigned_partitions, max_assigned_partitions y avg_assigned_partitions. Para obtener más información, consulte SVL_S3PARTITION_SUMMARY. |
| SVL_S3QUERY | Columnas: external_table_name, file_format, s3_scanned_rows, s3_scanned_bytes, s3query_returned_rows y s3query_returned_bytes. Para obtener más información, consulte SVL_S3QUERY. |
| SVL_S3QUERY_SUMMARY | Columnas: external_table_name, file_format, s3_scanned_rows, s3_scanned_bytes, s3query_returned_rows y s3query_returned_bytes. Para obtener más información, consulte SVL_S3QUERY_SUMMARY. |
| SVL_S3RETRIES | Columnas: file_size, location y message. Para obtener más información, consulte SVL_S3RETRIES. |
| SVL_SPECTRUM_SCAN_ERROR | Columnas: location, rowid, colname, original_value y modified_value. Para obtener más información, consulte SVL_SPECTRUM_SCAN_ERROR. |
| SVL_STATEMENTTEXT | Columnas: type y text. Para obtener más información, consulte SVL_STATEMENTTEXT. |
| SVL_STORED_PROC_CALL | Columnas: querytxt. Para obtener más información, consulte SVL_STORED_PROC_CALL. |
| SVL_STORED_PROC_MESSAGES | Columnas: querytext. Para obtener más información, consulte SVL_STORED_PROC_MESSAGES. |

| Tabla del sistema | Columnas confidenciales |
|---|---|
| SVL_UDF_LOG | Columnas: funcname. Para obtener más información, consulte SVL_UDF_LOG. |
| SVV_DISKUSAGE | Columnas: name, col, tbl, blocknum, num_values, minvalue y maxvalue. Para obtener más información, consulte SVV_DISKUSAGE. |
| SVV_QUERY_STATE | Columnas: rows, bytes y label. Para obtener más información, consulte SVV_QUERY_STATE. |
| SVV_TABLE_INFO | Columnas: table_id y table. Para obtener más información, consulte SVV_TABLE_INFO. |
| SVV_TRANSACTIONS | Columnas: relation. Para obtener más información, consulte SVV_TRANSACTIONS. |

## Consideraciones y limitaciones

A continuación se presentan las consideraciones y limitaciones que se aplican a las vistas creadas en el catálogo de datos.

- Las vistas de AWS Glue Data Catalog solo se admiten en clústeres aprovisionados de RA3 o grupos de trabajo de Redshift sin servidor.
- No puede crear una vista del catálogo de datos que se base en otra vista.
- Solo puede disponer de 10 tablas básicas en una vista de catálogo de datos.
- El definidor de la vista debe tener todos los permisos `SELECT GRANTABLE` en las tablas básicas.
- Las vistas solo pueden contener objetos y elementos integrados de Lake Formation. Los siguientes objetos no están permitidos dentro de una vista.
  - Tablas del sistema
  - Funciones definidas por el usuario (UDF)
  - Tablas, vistas, vistas materializadas y vistas de enlace de tiempo de ejecución de Redshift que no comparten datos administrados en Lake Formation.
- Las vistas no pueden contener tablas de Redshift Spectrum anidadas.

- Las representaciones de AWS Glue de los objetos básicos de una vista deben estar en la misma Cuenta de AWS y región que la vista.

# Consulta de datos espaciales en Amazon Redshift

Los datos espaciales describen la posición y la forma de una geometría en un espacio definido (un sistema de referencia espacial). Amazon Redshift admite los datos espaciales con el tipo de datos `GEOMETRY` y `GEOGRAPHY`, que contienen datos espaciales y, de forma opcional, el identificador de sistema de referencia espacial (SRID) de los datos.

Los datos espaciales contienen datos geométricos que se pueden usar para representar características geométricas. Este tipo de datos puede corresponder a informes meteorológicos, direcciones de mapas, tweets con posiciones geográficas, ubicaciones de almacenes y rutas de líneas aéreas. Los datos espaciales desempeñan un papel importante en el análisis, los informes y las previsiones de negocio.

Puede consultar los datos espaciales con las funciones SQL de Amazon Redshift. Los datos espaciales contienen valores geométricos de un objeto.

Las operaciones de tipo de datos `GEOMETRY` funcionan en el plano cartesiano. Aunque el identificador del sistema de referencia espacial (SRID) se almacena dentro del objeto, este SRID es simplemente un identificador del sistema de coordenadas y no desempeña ningún papel en los algoritmos utilizados para procesar los objetos `GEOMETRY`. Por el contrario, las operaciones en el tipo de datos `GEOGRAPHY` tratan las coordenadas dentro de los objetos como coordenadas esféricas de un esferoide. Este esferoide se define por el SRID, que hace referencia a un sistema de referencia espacial geográfica. De forma predeterminada, los tipos de datos `GEOGRAPHY` se crean con referencia espacial (SRID) 4326, y hacen referencia al Sistema Geodésico Mundial (WGS) 84. Para obtener más información acerca de SRID, consulte Sistema de referencia espacial en Wikipedia.

Puede utilizar la función ST_Transform para transformar las coordenadas de varios sistemas de referencia espacial. Una vez realizada la transformación de las coordenadas, también puede utilizar una conversión sencilla entre ambas, siempre que la entrada `GEOMETRY` esté codificada con el SRID geográfico. Esta conversión simplemente copia las coordenadas sin realizar ninguna transformación adicional. Por ejemplo:

```
SELECT ST_AsEWKT(ST_GeomFromEWKT('SRID=4326;POINT(10 20)')::geography);
```

```
st_asewkt
------------------------
```

```
SRID=4326;POINT(10 20)
```

Para entender mejor la diferencia entre los tipos de datos GEOMETRY y GEOGRAPHY, considere calcular la distancia entre el aeropuerto de Berlín (BER) y el aeropuerto de San Francisco (SFO) utilizando el Sistema Geodésico Mundial (WGS) 84. Si utiliza el tipo de datos GEOGRAPHY, el resultado se expresa en metros. Cuando se utiliza el tipo de datos GEOMETRY con SRID 4326, el resultado se expresa en grados y no se pueden convertir en metros porque la distancia de un grado depende de dónde se encuentren las geometrías del globo.

Los cálculos sobre los tipos de datos GEOGRAPHY se utilizan principalmente para cálculos realistas de tierra redonda, como el área precisa de un país sin distorsión. Pero calcularlo significa un costo mucho mayor. Por lo tanto, ST_Transform puede transformar las coordenadas en un sistema apropiado local de coordenadas proyectadas y realizar el cálculo en el tipo de datos GEOMETRY de forma más rápida.

Al usar datos espaciales puede ejecutar consultas para hacer lo siguiente:

- Encontrar la distancia entre dos puntos.
- Comprobar si un área (polígono) contiene otra área.
- Comprobar si una linestring se cruza con otra linestring o polígono.

Puede utilizar el tipo de datos GEOMETRY para almacenar los valores de datos espaciales. Un valor GEOMETRY en Amazon Redshift puede definir tipos de datos primitivos de geometría bidimensional (2D), tridimensional (3DZ), bidimensional con una medida (3DM) y cuatridimensional (4D):

- Una geometría bidimensional (2D) se especifica mediante dos coordenadas cartesianas (x, y) en un plano.
- Una geometría tridimensional (3DZ) se especifica mediante tres coordenadas cartesianas (x, y, z) en el espacio.
- Una geometría bidimensional con una medida (3DM) se especifica mediante tres coordenadas (x, y, m), donde las dos primeras son coordenadas cartesianas en un plano y la tercera es una medida.
- Una geometría cuatridimensional (4D) se especifica mediante cuatro coordenadas (x, y, z, m), donde las tres primeras son coordenadas cartesianas en un espacio y la tercera es una medida.

Para obtener más información acerca de los tipos de datos primitivos de geometría, consulte Well Known Text en Wikipedia.

Puede utilizar el tipo de datos GEOGRAPHY para almacenar los valores de datos espaciales. Un valor GEOGRAPHY en Amazon Redshift puede definir tipos de datos primitivos de geometría bidimensional (2D), tridimensional (3DZ), bidimensional con una medida (3DM) y cuatridimensional (4D):

- Una geometría bidimensional (2D) se especifica mediante coordenadas de longitud y latitud en un esferoide.
- Una geometría tridimensional (3DZ) se especifica mediante coordenadas de longitud, latitud y altitud en un esferoide.
- Una geometría bidimensional con una medida (3DM) se especifica mediante tres coordenadas (longitud, latitud, medida), donde las dos primeras son coordenadas angulares en una esfera y la tercera es una medida.
- Una geometría de cuatro dimensiones (4D) se especifica mediante cuatro coordenadas (longitud, latitud, altitud, medida), donde las tres primeras son longitud, latitud y altitud, y la cuarta es una medida.

Para obtener más información acerca de los sistemas de coordenadas geográficas, consulte Geographic coordinate system y Spherical coordinate system en Wikipedia.

Los tipos de datos GEOMETRY y GEOGRAPHY tienen los siguientes subtipos:

- POINT
- LINESTRING
- POLYGON
- MULTIPOINT
- MULTILINESTRING
- MULTIPOLYGON
- GEOMETRYCOLLECTION

Hay funciones SQL de Amazon Redshift que son compatibles con las siguientes representaciones de datos geométricos:

- GeoJSON
- Well-known Text (WKT)
- Extended well-known text (EWKT)
- Representación well-known binary (WKB)

- Extended well-known binary (EWKB)

Puede convertir entre los tipos de datos GEOMETRY y GEOGRAPHY.

El siguiente SQL convierte una cadena de líneas desde un GEOMETRY a un GEOGRAPHY.

```
SELECT ST_AsEWKT(ST_GeomFromText('LINESTRING(110 40, 2 3, -10 80, -7 9)')::geography);
```

```
 st_asewkt
--------------------------------------------
 SRID=4326;LINESTRING(110 40,2 3,-10 80,-7 9)
```

El siguiente SQL convierte una cadena de líneas desde un GEOGRAPHY a un GEOMETRY.

```
SELECT ST_AsEWKT(ST_GeogFromText('LINESTRING(110 40, 2 3, -10 80, -7 9)')::geometry);
```

```
 st_asewkt
--------------------------------------------
 SRID=4326;LINESTRING(110 40,2 3,-10 80,-7 9)
```

Amazon Redshift proporciona muchas funciones SQL para consultar los datos espaciales. A excepción de la función ST_IsValid, las funciones espaciales que aceptan un objeto GEOMETRY como argumento esperan que este objeto GEOMETRY sea una geometría válida. Si el objeto GEOMETRY o GEOGRAPHY no es válido, entonces el comportamiento de la función espacial no está definido. Para obtener más información acerca de la validez, consulte Validez geométrica.

Para obtener más detalles acerca de las funciones SQL para datos espaciales, consulte Funciones espaciales.

Para obtener más información acerca de cómo cargar los datos espaciales, consulte Carga de una columna de tipo de datos GEOMETRY o GEOGRAPHY.

Temas

- Tutorial: Uso de funciones SQL espaciales con Amazon Redshift
- Carga de un shapefile en Amazon Redshift
- Terminología de los datos espaciales de Amazon Redshift

- Consideraciones al usar datos espaciales con Amazon Redshift

# Tutorial: Uso de funciones SQL espaciales con Amazon Redshift

En este tutorial, se muestra cómo utilizar algunas de las funciones SQL espaciales con Amazon Redshift.

Para ello, debe consultar dos tablas utilizando funciones SQL espaciales. En el tutorial, se utilizan datos de conjuntos de datos públicos que relacionan los datos de ubicación de los alojamientos de alquiler con los códigos postales en Berlín, Alemania.

Temas


- Requisitos previos
- Paso 1: Crear tablas y cargar datos de prueba
- Paso 2: Consultar los datos espaciales
- Paso 3: Eliminar los recursos


## Requisitos previos

Para este tutorial, necesita los siguientes recursos:

- Un clúster y una base de datos de Amazon Redshift existentes a los que pueda acceder y actualizar. En el clúster existente, puede crear tablas, cargar datos de muestra y ejecutar consultas SQL para mostrar las funciones espaciales. El clúster debe tener dos nodos como mínimo. Para obtener información acerca de cómo crear un clúster, siga los pasos que se indican en la Guía de introducción a Amazon Redshift.
- Para utilizar el editor de consultas de Amazon Redshift, asegúrese de que el clúster se encuentra en una región de AWS que sea compatible con el editor de consultas. Para obtener más información, consulte Consulta de una base de datos mediante el editor de consultas en la Guía de administración de Amazon Redshift.
- Credenciales de AWS para su clúster de Amazon Redshift que le permiten cargar datos de prueba desde Amazon S3. Para obtener información sobre cómo acceder a otros servicios de AWS, como Amazon S3, consulte Autorización a Amazon Redshift para obtener acceso a los servicios de AWS.
- El rol de AWS Identity and Access Management (IAM) denominado `mySpatialDemoRole`, que tiene la política administrada `AmazonS3ReadOnlyAccess` adjuntada para leer los datos de

Amazon S3. Para crear un rol que tenga permiso para cargar datos de un bucket de Amazon S3, consulte Autorización de las operaciones COPY, UNLOAD y CREATE EXTERNAL SCHEMA mediante los roles de IAM en la Guía de administración de Amazon Redshift.

- Después de crear el rol de IAM `mySpatialDemoRole`, este necesita asociarse con su clúster de Amazon Redshift. Para obtener más información acerca de cómo crear esa asociación, consulte Autorización de las operaciones COPY, UNLOAD y CREATE EXTERNAL SCHEMA mediante los roles de IAM en la Guía de administración de Amazon Redshift.

## Paso 1: Crear tablas y cargar datos de prueba

Los datos de origen que se utilizaron en este se encuentran en archivos denominados `accommodations.csv` y `zipcodes.csv`.

El archivo `accommodations.csv` consiste en datos de código abierto de insideairbnb.com. El archivo `zipcodes.csv` proporciona códigos postales que son datos de código abierto del Instituto Nacional de Estadística de Berlín-Brandeburgo en Alemania (Amt für Statistik Berlín-Brandenburg). Ambos orígenes de datos se proporcionan bajo una licencia de Creative Commons. Los datos se limitan a la región de Berlín, Alemania. Estos archivos se encuentran en un bucket público de Amazon S3 para utilizarse en este tutorial.

De manera opcional, puede descargar los datos de origen desde los siguientes enlaces de Amazon S3:

- Datos de origen para la tabla `accommodations`.
- Datos de origen para la tabla `zipcode`.

Utilice el siguiente procedimiento para crear tablas y cargar datos de prueba.

Para crear tablas y cargar datos de prueba

1. Abra el editor de consultas de Amazon Redshift. Para obtener más información acerca de cómo trabajar con el editor de consultas, visite Consulta de una base de datos mediante el editor de consultas en la Guía de administración de Amazon Redshift.
2. Elimine todas las tablas utilizadas en este tutorial si ya existen en la base de datos. Para obtener más información, consulte Paso 3: Eliminar los recursos.
3. Cree la tabla `accommodations` para almacenar la ubicación geográfica de cada alojamiento (longitud y latitud), el nombre del listado y otros datos comerciales.

En este tutorial, se analiza el alquiler de habitaciones en Berlín, Alemania. En la columna `shape`, se almacenan los puntos geográficos de la ubicación de los alojamientos. Las otras columnas contienen información sobre el alquiler.

Para crear la tabla `accommodations`, ejecute la siguiente instrucción SQL en el editor de consultas de Amazon Redshift.

```
CREATE TABLE public.accommodations (
  id INTEGER PRIMARY KEY,
  shape GEOMETRY,
  name VARCHAR(100),
  host_name VARCHAR(100),
  neighbourhood_group VARCHAR(100),
  neighbourhood VARCHAR(100),
  room_type VARCHAR(100),
  price SMALLINT,
  minimum_nights SMALLINT,
  number_of_reviews SMALLINT,
  last_review DATE,
  reviews_per_month NUMERIC(8,2),
  calculated_host_listings_count SMALLINT,
  availability_365 SMALLINT
);
```

4. Cree la tabla `zipcode` en el editor de consultas para almacenar los códigos postales de Berlín.

Un código postal se define como un polígono en la columna `wkb_geometry`. En el resto de las columnas, se describen metadatos espaciales adicionales sobre el código postal.

Para crear la tabla `zipcode`, ejecute la siguiente instrucción SQL en el editor de consultas de Amazon Redshift.

```
CREATE TABLE public.zipcode (
  ogc_field INTEGER PRIMARY KEY NOT NULL,
  wkb_geometry GEOMETRY,
  gml_id VARCHAR(256),
  spatial_name VARCHAR(256),
  spatial_alias VARCHAR(256),
  spatial_type VARCHAR(256)
 );
```

5. Cargue las tablas con datos de muestra.

   Los datos de muestra para este tutorial se encuentran en un bucket de Amazon S3 que permite el acceso de lectura a todos los usuarios de AWS autenticados. Asegúrese de proporcionar credenciales de AWS válidas que permiten el acceso a Amazon S3.

   Para cargar datos de prueba en las tablas, ejecute los siguientes comandos COPY. Sustituya *account-number* por su número de cuenta de AWS. El segmento de la cadena de credenciales que está delimitado entre comillas simples no debe contener espacios ni saltos de línea.

```
COPY public.accommodations
FROM 's3://redshift-downloads/spatial-data/accommodations.csv'
DELIMITER ';'
IGNOREHEADER 1 REGION 'us-east-1'
CREDENTIALS 'aws_iam_role=arn:aws:iam::account-number:role/mySpatialDemoRole';
```

```
COPY public.zipcode
FROM 's3://redshift-downloads/spatial-data/zipcode.csv'
DELIMITER ';'
IGNOREHEADER 1 REGION 'us-east-1'
CREDENTIALS 'aws_iam_role=arn:aws:iam::account-number:role/mySpatialDemoRole';
```

6. Ejecute los siguientes comandos para comprobar que cada tabla se haya cargado de forma correcta.

```
select count(*) from accommodations;
```

```
select count(*) from zipcode;
```

   Los siguientes resultados muestran la cantidad de filas en cada tabla de los datos de prueba.

| Nombre de la tabla | Filas |
|---|---|
| accomodations | 22 248 |
| zipcode | 190 |

# Paso 2: Consultar los datos espaciales

Después de crear y cargar las tablas, puede consultarlas mediante las instrucciones SQL SELECT. Las siguientes consultas muestran parte de la información que se puede recuperar. Puede escribir muchas otras consultas que utilizan funciones espaciales para satisfacer sus necesidades.

Para consultar los datos espaciales

1. Realice una consulta para obtener el recuento de la cantidad total de listados almacenados en la tabla `accommodations`, como se muestra a continuación. El sistema de referencia espacial es el sistema geodésico mundial (WGS) 84, que tiene el identificador de referencia espacial único 4326.

```
SELECT count(*) FROM public.accommodations WHERE ST_SRID(shape) = 4326;
```

```
 count
-------
 22248
```

2. Recupere los objetos de geometría en formato de texto conocido (WKT) con algunos atributos adicionales. Además, puede comprobar si estos datos de código postal también se almacenan con el sistema geodésico mundial (WGS) 84, que utiliza el ID de referencia espacial (SRID) 4326. Los datos espaciales deben almacenarse en el mismo sistema de referencia espacial para que sean interoperables.

```
SELECT ogc_field, spatial_name, spatial_type, ST_SRID(wkb_geometry),
 ST_AsText(wkb_geometry)
FROM public.zipcode
ORDER BY spatial_name;
```

```
ogc_field  spatial_name  spatial_type  st_srid  st_astext
-----------------------------------------------------------------
0          10115        Polygon        4326     POLYGON((...))
4          10117        Polygon        4326     POLYGON((...))
8          10119        Polygon        4326     POLYGON((...))
...
(190 rows returned)
```

3. Seleccione el polígono de Berlin-Mitte (10117), un distrito de Berlín, en formato GeoJSON, su dimensión y la cantidad de puntos en este polígono.

```
SELECT ogc_field, spatial_name, ST_AsGeoJSON(wkb_geometry),
 ST_Dimension(wkb_geometry), ST_NPoints(wkb_geometry)
FROM public.zipcode
WHERE spatial_name='10117';
```

```
ogc_field  spatial_name  spatial_type
 st_dimension  st_npoint
-------------------------------------------------------------------------------------------
4           10117         {"type":"Polygon", "coordinates":[[[...]]]}    2
    331
```

4. Ejecute el siguiente comando SQL para ver cuántos alojamientos hay a menos de 500 metros de la Puerta de Brandenburgo.

```
SELECT count(*)
FROM public.accommodations
WHERE ST_DistanceSphere(shape, ST_GeomFromText('POINT(13.377704 52.516431)', 4326))
 < 500;
```

```
count
------
  29
```

5. Obtenga la ubicación aproximada de la Puerta de Brandenburgo a partir de los datos almacenados en los alojamientos que se mencionan como cercanos mediante la ejecución de la siguiente consulta.

   Esta consulta requiere una subselección. Esto implica un recuento diferente, ya que la ubicación solicitada no es la misma que la consulta anterior porque está más cerca de los alojamientos.

```
WITH poi(loc) as (
  SELECT st_astext(shape) FROM accommodations WHERE name LIKE '%brandenburg gate%'
)
SELECT count(*)
FROM accommodations a, poi p
```

```
WHERE ST_DistanceSphere(a.shape, ST_GeomFromText(p.loc, 4326)) < 500;
```

```
count
------
  60
```

6. Ejecute la siguiente consulta para mostrar los detalles de todos los alojamientos en los alrededores de la Puerta de Brandenburgo, ordenados por precio en orden descendente.

```
SELECT name, price, ST_AsText(shape)
FROM public.accommodations
WHERE ST_DistanceSphere(shape, ST_GeomFromText('POINT(13.377704 52.516431)', 4326))
 < 500
ORDER BY price DESC;
```

```
name                                                price  st_astext
-------------------------------------------------------------------------------------------
DUPLEX APARTMENT/PENTHOUSE in 5* LOCATION! 7583        300
 POINT(13.3826510209548 52.5159819722552)
DUPLEX-PENTHOUSE IN FIRST LOCATION! 7582               300
 POINT(13.3799997083855 52.5135918444834)
...
(29 rows returned)
```

7. Ejecute la siguiente consulta para recuperar el alojamiento más costoso con su código postal.

```
SELECT
  a.price, a.name, ST_AsText(a.shape),
  z.spatial_name, ST_AsText(z.wkb_geometry)
FROM accommodations a, zipcode z
WHERE price = 9000 AND ST_Within(a.shape, z.wkb_geometry);
```

```
price   name                                st_astext
     spatial_name      st_astext
-------------------------------------------------------------------------------------------
```

```
9000    Ueber den Dächern Berlins Zentrum    POINT(13.334436985013
 52.4979779501538)    10777             POLYGON((13.3318284987227
 52.4956021172799,...
```

8. Calcule el precio máximo, mínimo o medio de los alojamientos con una subconsulta.

   Con la siguiente consulta, se muestra el precio medio de los alojamientos por código postal.

```
SELECT
  a.price, a.name, ST_AsText(a.shape),
  z.spatial_name, ST_AsText(z.wkb_geometry)
FROM accommodations a, zipcode z
WHERE
  ST_Within(a.shape, z.wkb_geometry) AND
  price = (SELECT median(price) FROM accommodations)
ORDER BY a.price;
```

```
price name                            st_astext
 spatial_name   st_astext
-------------------------------------------------------------------------------------------

45    "Cozy room Berlin-Mitte"        POINT(13.3864349535358 52.5292016386514)
 10115          POLYGON((13.3658598465795 52.535659581048,...
...
(723 rows returned)
```

9. Ejecute la siguiente consulta para recuperar la cantidad de alojamientos que se muestran para Berlín. Los puntos calientes se agrupan por código postal y se ordenan por la cantidad de suministro, por lo que pueden encontrarse de esta manera.

```
SELECT z.spatial_name as zip, count(*) as numAccommodations
FROM public.accommodations a, public.zipcode z
WHERE ST_Within(a.shape, z.wkb_geometry)
GROUP BY zip
ORDER BY numAccommodations DESC;
```

```
zip   numaccommodations
---------------------------
10245 872
10247 832
```

```
10437 733
10115 664
...
(187 rows returned)
```

## Paso 3: Eliminar los recursos

El clúster seguirá acumulando cargos mientras esté en ejecución. Una vez que haya completado este tutorial, puede eliminar el clúster de muestra.

Si desea mantener el clúster, pero quiere recuperar el almacenamiento utilizado por las tablas de datos de prueba, ejecute los siguientes comandos para eliminar las tablas.

```
drop table public.accommodations cascade;
```

```
drop table public.zipcode cascade;
```

# Carga de un shapefile en Amazon Redshift

Puede utilizar el comando COPY para capturar los ESRI Shapefiles almacenados en Amazon S3 en las tablas de Amazon Redshift. Un shapefile almacena la ubicación geométrica y la información de los atributos de las características geográficas en un formato vectorial. Con el formato shapefile, se pueden describir de forma espacial distintos objetos espaciales como puntos, líneas y polígonos. Para obtener más información acerca de un shapefile, consulte Shapefile en Wikipedia.

El comando COPY admite el parámetro de formato de datos `SHAPEFILE`. De forma predeterminada, la primera columna del shapefile es una columna `GEOMETRY` o `IDENTITY`. Todas las columnas posteriores tienen el orden especificado en el shapefile. No obstante, la tabla de destino no necesita tener este mismo diseño porque se puede usar el mapeo de columnas de COPY para establecer el orden. Para obtener información acerca de la compatibilidad del formato shapefile del comando COPY, consulte SHAPEFILE.

En algunos casos, el tamaño de la geometría resultante puede ser mayor que el máximo para almacenar una geometría en Amazon Redshift. Si esto sucede, puede usar la opción `SIMPLIFY` o `SIMPLIFY AUTO` de COPY para simplificar las geometrías durante la ingesta de la siguiente manera:

- Especifique `SIMPLIFY` *`tolerance`* para simplificar todas las geometrías durante la ingesta utilizando el algoritmo de Ramer-Douglas-Peucker y la tolerancia dada.
- Especifique `SIMPLIFY AUTO` sin tolerancia para simplificar solo las geometrías mayores que el tamaño máximo mediante el algoritmo de Ramer-Douglas-Peucker. Con este enfoque, se calcula la tolerancia mínima, que es lo suficientemente grande como para almacenar el objeto dentro del límite de tamaño máximo.
- Especifique `SIMPLIFY AUTO` *`max_tolerance`* para simplificar solo las geometrías mayores que el tamaño máximo mediante el algoritmo de Ramer-Douglas-Peucker y la tolerancia calculada de forma automática. Este enfoque garantiza que la tolerancia no supere su valor máximo.

Para obtener información acerca del tamaño máximo de un valor de datos `GEOMETRY`, consulte Consideraciones al usar datos espaciales con Amazon Redshift.

En algunos casos, la tolerancia es lo suficientemente baja como para que el registro no pueda reducirse por debajo del tamaño máximo de un valor de datos `GEOMETRY`. En estos casos, puede utilizar la opción `MAXERROR` del comando COPY para ignorar todos los errores de ingesta o hasta un cierto número de ellos.

El comando COPY también admite la carga de shapefiles GZIP. Para ello, especifique el parámetro COPY GZIP. Con esta opción, todos los componentes del shapefile deben comprimirse de forma independiente y compartir el mismo sufijo de compresión.

Si existe un archivo de descripción de proyección (.prj) con el shapefile, Redshift lo utiliza para determinar el id. del sistema de referencia espacial (SRID). Si el SRID es válido, la geometría resultante tiene este SRID asignado. Si el valor SRID asociado a la geometría de entrada no existe, la geometría resultante tendrá cero como valor SRID. Puede desactivar la detección automática del ID del sistema de referencia espacial en el nivel de la sesión utilizando `SET read_srid_on_shapefile_ingestion` en `OFF`.

Consulte las vistas del sistema `SYS_SPATIAL_SIMPLIFY` o `SVL_SPATIAL_SIMPLIFY` para ver qué registros se han simplificado, junto con la tolerancia calculada. Cuando especifica `SIMPLIFY` *`tolerance`*, esta vista contiene un registro para cada operación COPY. De lo contrario, contiene un registro para cada geometría simplificada. Para obtener más información, consulte SYS_SPATIAL_SIMPLIFY o SVL_SPATIAL_SIMPLIFY.

Para ver ejemplos de carga de un shapefile, consulte Carga de un shapefile en Amazon Redshift.

# Terminología de los datos espaciales de Amazon Redshift

Los siguientes términos se utilizan para describir algunas funciones espaciales de Amazon Redshift.

## Cuadro delimitador

Un cuadro delimitador de una geometría o geografía se define como el producto vectorial (entre las dimensiones) de la extensión de las coordenadas de todos los puntos de la geometría o la geografía. En el caso de las geometrías bidimensionales, el cuadro delimitador es un rectángulo que incluye por completo todos los puntos de la geometría. Por ejemplo, un cuadro delimitador del polígono `POLYGON((0 0,1 0,0 2,0 0))` es el rectángulo definido por los puntos (0, 0) y (1, 2) como sus esquinas inferior izquierda y superior derecha. Amazon Redshift calcula previamente y almacena un cuadro delimitador dentro de una geometría para acelerar los predicados geométricos y las uniones espaciales. Por ejemplo, si los cuadros delimitadores de dos geometrías no forman una intersección, estas dos geometrías no pueden formar una intersección y no pueden ser parte del conjunto de resultados de una unión espacial que utiliza el predicado ST_Intersects.

Puede usar funciones espaciales para agregar (AddBBox), eliminar (DropBBox) y determinar el soporte (SupportsBBox) para un cuadro delimitador. Amazon Redshift admite el cálculo previo de los cuadros delimitadores para todos los subtipos de geometría.

En el siguiente ejemplo, se muestra cómo actualizar las geometrías existentes en una tabla para almacenarlas con un cuadro delimitador. Si la versión del clúster es la 1.0.26809 o una posterior, todas las geometrías nuevas se crean con un cuadro delimitador calculado previamente de forma predeterminada.

```
UPDATE my_table SET geom = AddBBox(geom) WHERE SupportsBBox(geom) = false;
```

Después de actualizar las geometrías existentes, se recomienda ejecutar el comando VACUUM en la tabla actualizada. Para obtener más información, consulte VACUUM.

Para establecer si las geometrías están codificadas con un cuadro delimitador durante una sesión, consulte default_geometry_encoding.

## Validez geométrica

Los algoritmos geométricos que se utilizan en Amazon Redshift suponen que la geometría de entrada es una geometría válida. Si la entrada de un algoritmo no es válida, entonces el resultado no

está definido. En la siguiente sección, se describen las definiciones de validez geométrica utilizadas en Amazon Redshift para cada subtipo de geometría.

Point (Punto)

Un punto se considera válido si se cumple alguna de las siguientes condiciones:

- El punto es el punto vacío.
- Todas las coordenadas de un punto son números finitos de coma flotante.

Un punto puede ser el punto vacío.

Linestring (Cadena de línea)

Una cadena de línea se considera válida si se cumple alguna de las siguientes condiciones:

- La cadena de línea está vacía; es decir, no contiene puntos.
- Todos los puntos de una cadena de línea no vacía tienen coordenadas que son números finitos de coma flotante.
- La cadena de línea, si no está vacía, debe ser unidimensional; es decir, no se puede degenerar hasta convertirse en un punto.

Una cadena de línea no puede contener puntos vacíos.

Una cadena de línea puede tener puntos consecutivos duplicados.

Una cadena de línea puede tener autointersecciones.

Polygon (Polígono)

Un polígono se considera válido si se cumple alguna de las siguientes condiciones:

- El polígono está vacío; es decir, no contiene puntos.
- Si no está vacío, un polígono es válido si se cumplen todas las siguientes condiciones:
  - Todos los anillos del polígono son válidos. Un anillo se considera válido si se cumplen todas las siguientes condiciones:
    - Todos los puntos del anillo tienen coordenadas que son números finitos de coma flotante.
    - El anillo está cerrado; es decir, su primer punto y su último punto coinciden.
    - El anillo no tiene autointersecciones.
    - El anillo es bidimensional.
  - Las orientaciones de los anillos del polígono deben ser coherentes. Es decir, si recorre cualquier anillo, el interior del polígono está tanto a su derecha como a su izquierda.

- Todos los anillos interiores deben estar dentro del anillo exterior del polígono.
- Todos los anillos interiores deben tener la misma orientación y debe ser la orientación opuesta a la del anillo exterior.
- Los anillos interiores no se pueden anidar; es decir, un anillo interior no puede estar dentro de otro anillo interior.
- Las intersecciones de los anillos interiores y exteriores solo pueden tener lugar en un número finito de puntos.
- El interior del polígono debe estar conectado de forma sencilla.

Un polígono no puede contener puntos vacíos.

Multipoint (Multipunto)

Un multipunto se considera válido si se cumple alguna de las siguientes condiciones:

- El multipunto está vacío; es decir, no contiene puntos.
- Un multipunto no está vacío y todos los puntos son válidos según la definición de validez de los puntos.

Un multipunto puede contener uno o más puntos vacíos.

Un multipunto puede tener puntos duplicados.

Cadenas de líneas múltiples

Una cadena de varias líneas se considera válida si se cumple alguna de las siguientes condiciones:

- La cadena de varias líneas está vacía; es decir, no contiene puntos.
- Todas las cadenas de línea de una cadena de varias líneas no vacía son válidas según la definición de validez de las cadenas de línea.

Una cadena de varias líneas no vacía que consta solo de cadenas de línea vacías se considera válida.

Una cadena de línea vacía en una cadena de varias líneas no afecta su validez.

Una cadena de varias líneas puede tener cadenas de línea con puntos consecutivos duplicados.

Una cadena de varias líneas puede tener autointersecciones.

Una cadena de varias líneas no puede contener puntos vacíos.

Multipolygon (Multipolígono)

Un multipolígono se considera válido si se cumple alguna de las siguientes condiciones:

- El multipolígono no contiene ningún polígono (está vacío).
- El multipolígono no está vacío y se cumplen todas las siguientes condiciones:
  - Todos los polígonos del multipolígono son válidos.
  - No hay dos polígonos en el multipolígono que puedan formar una intersección en un número infinito de puntos. En particular, esto implica que el interior de dos polígonos cualesquiera no se puede formar una intersección y que solo se pueden tocar en un número finito de puntos.

Un polígono vacío en un multipolígono no invalida al multipolígono.

Un multipolígono no puede contener puntos vacíos.

Geometry collection (Colección de geometría)

Una colección de geometría se considera válida si se cumple alguna de las siguientes condiciones:

- La colección de geometría está vacía; es decir, no contiene geometrías.
- Todas las geometrías de una colección de geometría no vacía son válidas.

Esta definición sigue siendo válida, aunque de manera recursiva, para las colecciones de geometría anidadas.

Una colección de geometría puede contener puntos vacíos y multipuntos con puntos vacíos.

## Simplicidad geométrica

Los algoritmos geométricos que se utilizan en Amazon Redshift suponen que la geometría de entrada es una geometría válida. Si la entrada de un algoritmo no es válida, entonces la verificación de simplicidad no está definida. En la siguiente sección, se describen las definiciones de simplicidad geométrica utilizadas en Amazon Redshift para cada subtipo de geometría.

Point (Punto)

Un punto válido se considera simple si se cumple alguna de las siguientes condiciones:

- Un punto válido siempre se considera simple.
- Un punto vacío se considera simple.

Linestring (Cadena de línea)

Una cadena de línea válida se considera simple si se cumple alguna de las siguientes condiciones:

- La cadena de línea está vacía.
- La cadena de línea no está vacía y se cumplen todas las siguientes condiciones:
  - No tiene puntos consecutivos duplicados.
  - No tiene autointersecciones, excepto, posiblemente, por su primer punto y su último punto, que pueden coincidir. En otras palabras, la cadena de línea no puede tener autointersecciones, excepto en los puntos límite.

Polygon (Polígono)

Un polígono válido se considera simple si no contiene puntos consecutivos duplicados.

Multipoint (Multipunto)

Un multipunto válido se considera simple si se cumple alguna de las siguientes condiciones:

- El multipunto está vacío; es decir, no contiene puntos.
- No hay dos puntos no vacíos del multipunto que coincidan.

Cadenas de líneas múltiples

Una cadena de varias líneas válida se considera simple si se cumple alguna de las siguientes condiciones:

- La cadena de varias líneas está vacía.
- La cadena de varias líneas no está vacía y se cumplen todas las siguientes condiciones:
  - Todas sus cadenas de línea son simples.
  - Dos cadenas de línea cualesquiera de la cadena de varias líneas no forman una intersección, excepto en los puntos límites de las dos líneas.

Una cadena de varias líneas no vacía que consta solo de cadenas de línea vacías se considera vacía.

Una cadena de línea vacía en una cadena de varias líneas no afecta su simplicidad.

Una cadena de línea cerrada en una cadena de varias líneas no puede forman una intersección con ninguna otra cadena de línea en la cadena de varias líneas.

Una cadena de varias líneas no puede tener cadenas de línea con puntos consecutivos duplicados.

Multipolygon (Multipolígono)

Un multipolígono válido se considera simple si no contiene puntos consecutivos duplicados.

Geometry collection (Colección de geometría)

Una colección de geometría válida se considera simple si se cumple alguna de las siguientes condiciones:

- La colección de geometría está vacía; es decir, no contiene geometrías.
- Todas las geometrías de una colección de geometría no vacía son simples.

Esta definición sigue siendo válida, aunque de manera recursiva, para las colecciones de geometría anidadas.

## H3

El H3 es un sistema de cuadrícula de indexación geoespacial jerárquico que permite indexar las coordenadas espaciales con una resolución de metros cuadrados. Los datos indexados se pueden unir en conjuntos de datos dispares y agregarse con diferentes niveles de precisión. El H3 admite una variedad de algoritmos y optimizaciones basados en la cuadrícula, incluidos los vecinos más cercanos, la ruta más corta, el suavizado de gradientes y más. Los índices H3 hacen referencia a celdas que pueden ser hexágonos o pentágonos. El espacio se subdivide jerárquicamente de conformidad con una resolución. H3 admite 16 resoluciones, de 0 a 15, ambas incluidas. `0` es la más basta y `15` la más fina.

Amazon Redshift proporciona las siguientes funciones espaciales de H3:

- H3_Boundary
- H3_Center
- H3_FromLongLat
- H3_FromPoint
- H3_Polyfill

# Consideraciones al usar datos espaciales con Amazon Redshift

A continuación, se muestran las consideraciones que existen cuando se utilizan datos espaciales con Amazon Redshift:

- El tamaño máximo del objeto GEOMETRY o GEOGRAPHY es de 1 048 447 bytes.
- Amazon Redshift Spectrum no es compatible de forma nativa con los datos espaciales. Por lo tanto, no puede crear o cambiar una tabla externa con una columna GEOMETRY o GEOGRAPHY.
- Los tipos de datos para funciones definidas por el usuario Python (UDF) no admiten los tipos de datos GEOMETRY o GEOGRAPHY.
- No puede utilizar una columna GEOMETRY o GEOGRAPHY como clave de ordenación o clave de distribución para una tabla de Amazon Redshift.
- No puede utilizar las columnas GEOMETRY o GEOGRAPHY en cláusulas SQL ORDER BY, GROUP BY o DISTINCT.
- No puede utilizar las columnas GEOMETRY o GEOGRAPHY en varias funciones SQL.
- No puede realizar una operación UNLOAD en las columnas GEOMETRY o GEOGRAPHY en cada formato. Puede realizar una operación UNLOAD en las columnas GEOMETRY o GEOGRAPHY para archivos de texto o de valores separados por comas (CSV). Cuando lo haga, se escriben datos GEOMETRY o GEOGRAPHY en formato EWKB hexadecimal. Si el tamaño de los datos EWKB es mayor de 4 MB, se producirá una advertencia porque los datos no se podrán luego cargar en una tabla.
- La codificación de compresión admitida en datos GEOMETRY o GEOGRAPHY es RAW.
- Al usar controladores JDBC o ODBVC, utiliza mapeos de tipos personalizados. En este caso, la aplicación cliente debe tener información sobre qué parámetros de un objeto `ResultSet` son objetos GEOMETRY o GEOGRAPHY. La operación `ResultSetMetadata` devuelve tipo VARCHAR.
- Para copiar una fecha geográfica de un SHAPEFILE, primero incorpore una columna GEOMETRY y, a continuación, convierta los objetos a objetos GEOGRAPHY.

Las siguientes funciones no espaciales pueden aceptar una entrada de tipo GEOMETRY o GEOGRAPHY columnas de tipo GEOMETRY o GEOGRAPHY:

- La función de agregación COUNT
- Las expresiones condicionales COALESCE y NVL
- Expresiones CASE
- La codificación predeterminada para GEOMETRY y GEOGRAPHY es RAW. Para obtener más información, consulte Codificaciones de compresión.

# Consulta de datos con consultas federadas en Amazon Redshift

Con las consultas federadas en Amazon Redshift, puede consultar y analizar los datos en bases de datos operativas, almacenamientos de datos y lagos de datos. Gracias a la característica de consulta federada, puede integrar consultas desde Amazon Redshift sobre datos activos en bases de datos externas con consultas en los entornos de Amazon Redshift y Amazon S3. Las consultas federadas pueden utilizarse en bases de datos externas en Amazon RDS for PostgreSQL, la Edición compatible con PostgreSQL de Amazon Aurora, Amazon RDS for MySQL y la Edición compatible con MySQL de Amazon Aurora.

Puede utilizar consultas federadas para incorporar datos en vivo como parte de sus aplicaciones de inteligencia comercial (BI) y de informes. Por ejemplo, para facilitar la ingesta de datos en Amazon Redshift, puede utilizar consultas federadas para hacer lo siguiente:

- Consultar las bases de datos operativas directamente.
- Aplicar transformaciones rápidamente.
- Cargar datos en las tablas de destino sin necesidad de canalizaciones complejas de extracción, transformación y carga (ETL).

Con el objetivo de reducir el movimiento de los datos a través de la red y mejorar el rendimiento, Amazon Redshift distribuye parte del cómputo para consultas federadas directamente en las bases de datos operativas remotas. Además, Amazon Redshift utiliza su capacidad de procesamiento en paralelo para permitir la ejecución de estas consultas, cuando sea necesario.

Cuando se ejecutan consultas federadas, Amazon Redshift primero genera una conexión de cliente a la instancia de base de datos de clúster de RDS o Aurora DB desde el nodo principal para recuperar los metadatos de la tabla. Desde un nodo informático, Amazon Redshift emite subconsultas con un predicado que se delega y recupera las filas de resultados. Luego, Amazon Redshift distribuye las filas de resultados entre los nodos informáticos para continuar su procesamiento.

Se registran los detalles sobre las consultas enviadas a la base de datos de Amazon Aurora PostgreSQL o a la base de datos Amazon RDS for PostgreSQL en la vista SVL_FEDERATED_QUERY del sistema.

Temas

- Introducción al uso de consultas federadas en PostgreSQL
- Introducción al uso de consultas federadas en PostgreSQL con AWS CloudFormation
- Introducción al uso de consultas federadas en MySQL
- Creación de un secreto y rol de IAM para utilizar consultas federadas
- Ejemplos de uso de una consulta federada
- Diferencias de los tipos de datos entre Amazon Redshift y las bases de datos PostgreSQL y MySQL compatibles
- Consideraciones al obtener acceso a los datos federados con Amazon Redshift


# Introducción al uso de consultas federadas en PostgreSQL

Para crear una consulta federada, siga este enfoque general:

1. Configure la conectividad desde su clúster de Amazon Redshift a su instancia de base de datos de Amazon RDS o Aurora PostgreSQL.

   Para ello, asegúrese de que su instancia de base de datos de RDS PostgreSQL o Aurora PostgreSQL pueda aceptar conexiones desde el clúster de Amazon Redshift. Se recomienda que el clúster de Amazon Redshift y la instancia de Amazon RDS o Aurora PostgreSQL se encuentren en la misma nube virtual privada (VPC) y el mismo grupo de subredes. De esta manera, puede agregar el grupo de seguridad para el clúster de Amazon Redshift a las reglas de entrada del grupo de seguridad que corresponde a la instancia de base de datos de RDS o Aurora PostgreSQL.

   También puede configurar la interconexión de VPC u otras redes que permitan a Amazon Redshift establecer conexiones a su instancia de RDS o Aurora PostgreSQL. Para obtener más información acerca de las redes de VPC, consulte lo siguiente.

   - ¿Qué es una interconexión de VPC? en la Guía de interconexión de Amazon VPC
   - Uso de una instancia de base de datos en una VPC en la Guía del usuario de Amazon RDS

   > **Note**
   >
   > Hay casos en los que debe habilitar el enrutamiento de VPC mejorado: por ejemplo, si su clúster de Amazon Redshift se encuentra en una VPC diferente a la de su instancia de RDS o Aurora PostgreSQL, o si se encuentran en la misma VPC y sus rutas así lo

requieren. De lo contrario, podría recibir errores de tiempo de espera agotado al ejecutar una consulta federada.

2. Configure los secretos en AWS Secrets Manager para sus bases de datos de RDS PostgreSQL y Aurora PostgreSQL. A continuación, referencie los secretos en las políticas de acceso y los roles de AWS Identity and Access Management (IAM). Para obtener más información, consulte Creación de un secreto y rol de IAM para utilizar consultas federadas.

   > **Note**
   >
   > Si el clúster utiliza el enrutamiento de VPC mejorado, es posible que tenga configurar un punto de enlace de tipo interfaz de la VPC para AWS Secrets Manager. Esto es necesario cuando la VPC y la subred del clúster de Amazon Redshift no tienen acceso al punto de enlace público de AWS Secrets Manager. Cuando se utiliza un punto de enlace de tipo interfaz de la VPC, la comunicación entre el clúster de Amazon Redshift de la VPC y AWS Secrets Manager se dirige de forma privada desde la VPC hasta la interfaz del punto de enlace. Para más información, consulte Creación de un punto de conexión de interfaz en la Guía del usuario de Amazon VPC.

3. Aplique el rol de IAM creado anteriormente al clúster de Amazon Redshift. Para obtener más información, consulte Creación de un secreto y rol de IAM para utilizar consultas federadas.
4. Conéctese a sus bases de datos de RDS PostgreSQL y Aurora PostgreSQL con un esquema externo. Para obtener más información, consulte CREATE EXTERNAL SCHEMA. Para obtener ejemplos sobre cómo utilizar la consulta federada, consulte Ejemplos de uso de una consulta federada.
5. Ejecute las consultas SQL con una referencia al esquema externo que revisa las bases de datos de RDS PostgreSQL y Aurora PostgreSQL.

# Introducción al uso de consultas federadas en PostgreSQL con AWS CloudFormation

Puede utilizar consultas federadas para realizar consultas en bases de datos operativas. En esta guía de introducción, puede automatizar la configuración mediante una pila de AWS CloudFormation de muestra para habilitar una consulta federada desde un clúster de Amazon Redshift hacia una base de datos sin servidor de Aurora PostgreSQL. Puede ponerse en marcha rápidamente sin tener que ejecutar instrucciones SQL para aprovisionar sus recursos.

La pila crea un esquema externo que hace referencia a la instancia de Aurora PostgreSQL, que incluye tablas con datos de muestra. Puede consultar tablas del esquema externo desde el clúster de Redshift.

Si desea en cambio empezar a utilizar las consultas federadas ejecutando instrucciones SQL para configurar un esquema externo sin utilizar CloudFormation, consulte Introducción al uso de consultas federadas en PostgreSQL.

Antes de ejecutar la pila de CloudFormation para las consultas federadas, asegúrese de tener una base de datos sin servidor de la Edición compatible con PostgreSQL de Amazon Aurora con la API de datos activada. Puede activar la API de datos en las propiedades de la base de datos. Si no puede encontrar la configuración, verifique bien que está ejecutando una instancia sin servidor de Aurora PostgreSQL. Asegúrese también de tener un clúster de Amazon Redshift que utilice nodos RA3. Se recomienda que tanto el clúster de Redshift como la instancia sin servidor de Aurora PostgreSQL se encuentren en la misma nube virtual privada y el mismo grupo de subredes. De esta manera, puede agregar el grupo de seguridad para el clúster de Amazon Redshift a las reglas de entrada del grupo de seguridad de la instancia de base de datos de Aurora PostgreSQL.

Para obtener más información acerca de cómo empezar a configurar un clúster de Amazon Redshift, consulte Introducción a los almacenamientos de datos aprovisionados de Amazon Redshift. Para obtener más información acerca de cómo configurar recursos con CloudFormation, consulte ¿Qué es AWS CloudFormation? Para obtener más información acerca de cómo configurar una base de datos de clúster de Aurora DB, consulte Creación de un clúster de base de datos de clúster de Aurora DB sin servidor v1.

## Lanzamiento de una pila de CloudFormation para las consultas federadas de Redshift

Utilice el siguiente procedimiento para lanzar la pila de CloudFormation para Amazon Redshift con el objetivo de habilitar las consultas federadas. Antes de hacerlo, asegúrese de tener configurados el clúster de Amazon Redshift y la instancia sin servidor de Aurora PostgreSQL.

Para lanzar una pila de CloudFormation para las consultas federadas

1. Haga clic aquí en Launch CFN stack (Lanzar pila de CFN) para lanzar el servicio CloudFormation en la AWS Management Console.

   Si se le solicita, inicie sesión.

El proceso de creación de la pila se inicia si hace referencia a un archivo de plantilla de CloudFormation que se almacena en Amazon S3. Una plantilla de CloudFormation es un archivo de texto en formato JSON que establece los recursos de AWS que componen una pila.

2. Elija Next (Siguiente) para ingresar los detalles de la pila.
3. En Parameters (Parámetros), para el clúster, ingrese lo siguiente:

   - el nombre del clúster de Amazon Redshift, por ejemplo, **`ra3-consumer-cluster`**
   - un nombre de base de datos específico; por ejemplo, **`dev`**
   - el nombre de un usuario de base de datos; por ejemplo, **`consumeruser`**

   También ingrese los parámetros de la base de datos del clúster de Aurora DB, incluidos el usuario, el nombre de la base de datos, el puerto y el punto de conexión. Recomendamos utilizar un clúster de prueba y una base de datos sin servidor de prueba porque la pila crea varios objetos de base de datos.

   Elija Next (Siguiente).

   Aparecerán las opciones de la pila.
4. Elija Next (Siguiente) para aceptar la configuración predeterminada.
5. En Capabilities (Capacidades), elija I acknowledge that AWS CloudFormation might create IAM resources (Confirmo que AWS CloudFormation puede crear recursos de IAM).
6. Seleccione Crear pila.

Seleccione Crear pila. CloudFormation aprovisiona los recursos de la plantilla, proceso que demora unos 10 minutos, y crea un esquema externo.

Si se produce un error durante la creación de la pila, haga lo que se indica a continuación:

- Consulte la pestaña Events (Eventos) de CloudFormation para obtener información que pueda serle de utilidad para resolver el error.
- Asegúrese de haber ingresado el nombre correcto, el nombre de la base de datos y el nombre del usuario de la base de datos para el clúster de Redshift. Compruebe también los parámetros de la instancia de Aurora PostgreSQL.
- Asegúrese de que el clúster tenga nodos RA3.

- Asegúrese de que la base de datos y el clúster de Redshift estén en la misma subred y el mismo grupo de seguridad.

## Consulta de datos desde el esquema externo

Para utilizar el siguiente procedimiento, asegúrese de tener los permisos necesarios para ejecutar consultas en el clúster y la base de datos descritos.

Para consultar una base de datos externa con las consultas federadas

1. Conéctese a la base de datos de Redshift que ingresó cuando creó la pila por medio de una herramienta de cliente, como el editor de consultas de Redshift.
2. Consulta del esquema externo creado por la pila

   ```
   select * from svv_external_schemas;
   ```

   La vista SVV_EXTERNAL_SCHEMAS devuelve información sobre los esquemas externos disponibles. En este caso, se devuelve el esquema externo creado por la pila, `myfederated_schema`. Es posible que también se devuelvan otros esquemas externos si tiene alguno configurado. La vista también devuelve la base de datos asociada del esquema. La base de datos es la base de datos del clúster de Aurora DB que ingresó cuando creó la pila. La pila agrega una tabla a la base de datos del clúster de Aurora DB, llamada `category` y otra tabla llamada `sales`.
3. Ejecute consultas de SQL en las tablas del esquema externo que hace referencia a la base de datos de Aurora PostgreSQL. En el siguiente ejemplo, se muestra una consulta.

   ```
   SELECT count(*) FROM myfederated_schema.category;
   ```

   La tabla `category` devuelve varios registros. También puede devolver registros desde la tabla `sales`.

   ```
   SELECT count(*) FROM myfederated_schema.sales;
   ```

   Para obtener más ejemplos, consulte Ejemplos de uso de una consulta federada.

# Introducción al uso de consultas federadas en MySQL

Para crear una consulta federada en las bases de datos MySQL, siga el siguiente enfoque general:

1. Configure la conectividad desde su clúster de Amazon Redshift a la instancia de base de datos de Amazon RDS o Aurora MySQL.

   Para ello, asegúrese de que su instancia de base de datos de RDS MySQL o Aurora MySQL pueda aceptar conexiones desde el clúster de Amazon Redshift. Se recomienda que el clúster de Amazon Redshift y la instancia de Amazon RDS o Aurora MySQL se encuentren en la misma nube virtual privada (VPC) y el mismo grupo de subredes. De esta manera, puede agregar el grupo de seguridad para el clúster de Amazon Redshift a las reglas de entrada del grupo de seguridad para la instancia de base de datos de RDS o Aurora MySQL.

   También puede configurar la interconexión de VPC u otras redes que permitan a Amazon Redshift establecer conexiones con su instancia de RDS o Aurora MySQL. Para obtener más información acerca de las redes de VPC, consulte lo siguiente.

   - ¿Qué es una interconexión de VPC? en la Guía de interconexión de Amazon VPC
   - Uso de una instancia de base de datos en una VPC en la Guía del usuario de Amazon RDS

   > **Note**
   >
   > Si su clúster de Amazon Redshift se encuentra en una VPC diferente a la de la instancia de RDS o Aurora MySQL, habilite el enrutamiento de VPC mejorado. De lo contrario, podría recibir errores de tiempo de espera agotado al ejecutar una consulta federada.

2. Configure secretos en AWS Secrets Manager para sus bases de datos de RDS MySQL y Aurora MySQL. A continuación, referencie los secretos en las políticas de acceso y los roles de AWS Identity and Access Management (IAM). Para obtener más información, consulte Creación de un secreto y rol de IAM para utilizar consultas federadas.

   > **Note**
   >
   > Si el clúster utiliza el enrutamiento de VPC mejorado, es posible que tenga configurar un punto de enlace de tipo interfaz de la VPC para AWS Secrets Manager. Esto es necesario cuando la VPC y la subred del clúster de Amazon Redshift no tienen acceso al punto de enlace público de AWS Secrets Manager. Cuando se utiliza un punto de enlace de tipo interfaz de la VPC, la comunicación entre el clúster de Amazon Redshift de la VPC y AWS

Secrets Manager se dirige de forma privada desde la VPC hasta la interfaz del punto de enlace. Para más información, consulte Creación de un punto de conexión de interfaz en la Guía del usuario de Amazon VPC.

3. Aplique el rol de IAM creado anteriormente al clúster de Amazon Redshift. Para obtener más información, consulte Creación de un secreto y rol de IAM para utilizar consultas federadas.
4. Conéctese a sus bases de datos de RDS MySQL y Aurora MySQL con un esquema externo. Para obtener más información, consulte CREATE EXTERNAL SCHEMA. Para obtener ejemplos sobre cómo utilizar las consultas federadas, consulte Ejemplo de uso de una consulta federada con MySQL.
5. Ejecute las consultas SQL con una referencia al esquema externo que revisa las bases de datos de RDS MySQL y Aurora MySQL.

# Creación de un secreto y rol de IAM para utilizar consultas federadas

En los pasos siguientes se muestra cómo crear un secreto y un rol de IAM para utilizarlos con las consultas federadas.

## Requisitos previos

Asegúrese de cumplir los siguientes requisitos previos a fin de crear un secreto y un rol de IAM para utilizarlos con consultas federadas:

- una instancia de base de datos de RDS PostgreSQL, Aurora PostgreSQL, RDS MySQL o Aurora MySQL que cuente con autenticación de nombre de usuario y contraseña
- un clúster de Amazon Redshift con una versión de mantenimiento del clúster que admita las consultas federadas.

Para crear un secreto (nombre de usuario y contraseña) con AWS Secrets Manager

1. Inicie sesión en la consola de Secrets Manager con la cuenta dueña de la instancia del clúster de RDS o Aurora DB.
2. Elija Almacenar un secreto nuevo.
3. Elija el mosaico Credentials for RDS database (Credenciales para la base de datos RDS). En User name (Nombre de usuario) y Contraseña (Password), introduzca valores para la instancia.

Confirme o elija un valor para Encryption key (Clave de cifrado). A continuación, elija la base de datos de RDS a la que obtendrá acceso su secreto.

> **Note**
>
> Se recomienda utilizar la clave de cifrado predeterminada (`DefaultEncryptionKey`). Si utiliza una clave de cifrado personalizada, el rol de IAM que se utiliza para tener acceso al secreto debe añadirse como usuario clave.

4. Escriba un nombre para el secreto, continúe con los pasos de creación con las opciones predeterminadas y elija Store (Guardar).
5. Consulte su secreto y anote el valor de Secret ARN (ARN de secreto) que creó para identificar el secreto.

Para crear una directiva de seguridad mediante el secreto

1. Inicie sesión en AWS Management Console Management Console y abra la consola IAM en https://console.aws.amazon.com/iam/.
2. Cree una política con JSON similar a la siguiente.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AccessSecret",
            "Effect": "Allow",
            "Action": [
                "secretsmanager:GetResourcePolicy",
                "secretsmanager:GetSecretValue",
                "secretsmanager:DescribeSecret",
                "secretsmanager:ListSecretVersionIds"
            ],
            "Resource": "arn:aws:secretsmanager:us-west-2:123456789012:secret:my-
rds-secret-VNenFy"
        },
        {
            "Sid": "VisualEditor1",
            "Effect": "Allow",
            "Action": [
                "secretsmanager:GetRandomPassword",
```

```
                "secretsmanager:ListSecrets"
            ],
            "Resource": "*"
        }
    ]
}
```

Para recuperar el secreto, tiene que enumerar y leer las acciones. Se recomienda restringir el recurso al secreto específico que ha creado. Para ello, utilice el nombre de recurso de Amazon (ARN) del secreto para limitar el recurso. También puede especificar los permisos y recursos mediante el editor visual en la consola de IAM.

3. Asigne un nombre a la política y termine de crearla.
4. Vaya a IAM roles (Roles de IAM).
5. Cree un rol de IAM para Redshift - Customizable (Redshift - Personalizable).
6. Asocie la política de IAM que acaba de crear a un rol de IAM existente o cree uno nuevo y asóciele la política.
7. En la pestaña Trust relationships (Relaciones de confianza) de su rol de IAM, confirme que el rol contiene la entidad de confianza `redshift.amazonaws.com`.
8. Anote el valor de Role ARN (ARN de rol) que ha creado. Este ARN tiene acceso al secreto.

Para adjuntar el rol de IAM al clúster de Amazon Redshift

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. En el menú de navegación, elija Clusters (Clústeres). Se enumeran los clústeres de su cuenta en la región de AWS actual.
3. Seleccione el nombre del clúster en la lista para ver más detalles sobre un clúster.
4. En Actions (Acciones), seleccione Manage IAM roles (Administrar funciones de IAM). Aparecerá la página Manage IAM roles (Administrar los roles de IAM).
5. Añada su rol de IAM al clúster.

# Ejemplos de uso de una consulta federada

En los siguientes ejemplos, se muestra cómo ejecutar una consulta federada. Ejecute la SQL con su cliente SQL conectado a la base de datos de Amazon Redshift.

# Ejemplo de uso de una consulta federada con PostgreSQL

En el siguiente ejemplo, se muestra cómo configurar una consulta federada que referencie una base de datos de Amazon Redshift, una base de datos de Aurora PostgreSQL y Amazon S3. En este ejemplo, se muestra cómo funcionan las consultas federadas. Para ejecutarla en su propio entorno, cámbiela para que se ajuste a su entorno. Para obtener información sobre los requisitos previos para ello, consulte Introducción al uso de consultas federadas en PostgreSQL.

Cree un esquema externo que referencie una base de datos de Aurora PostgreSQL.

```
CREATE EXTERNAL SCHEMA apg
FROM POSTGRES
DATABASE 'database-1' SCHEMA 'myschema'
URI 'endpoint to aurora hostname'
IAM_ROLE 'arn:aws:iam::123456789012:role/Redshift-SecretsManager-RO'
SECRET_ARN 'arn:aws:secretsmanager:us-west-2:123456789012:secret:federation/test/
dataplane-apg-creds-YbVKQw';
```

Cree otro esquema externo que referencie Amazon S3, que usa Amazon Redshift Spectrum. Además, conceder permiso para utilizar el esquema a `public`.

```
CREATE EXTERNAL SCHEMA s3
FROM DATA CATALOG
DATABASE 'default' REGION 'us-west-2'
IAM_ROLE 'arn:aws:iam::123456789012:role/Redshift-S3';

GRANT USAGE ON SCHEMA s3 TO public;
```

Muestre el recuento de filas de la tabla de Amazon Redshift.

```
SELECT count(*) FROM public.lineitem;

  count
----------
25075099
```

Muestre el recuento de filas de la tabla de Aurora PostgreSQL.

```
SELECT count(*) FROM apg.lineitem;

count
```

```
-------
11760
```

Muestre el recuento de filas en Amazon S3.

```
SELECT count(*) FROM s3.lineitem_1t_part;

   count
------------
6144008876
```

Cree una vista de las tablas de Amazon Redshift, Aurora PostgreSQL y Amazon S3. Esta vista se utiliza para ejecutar las consultas federadas.

```
CREATE VIEW lineitem_all AS
  SELECT
 l_orderkey,l_partkey,l_suppkey,l_linenumber,l_quantity,l_extendedprice,l_discount,l_tax,l_retu
        l_shipdate::date,l_commitdate::date,l_receiptdate::date,
 l_shipinstruct ,l_shipmode,l_comment
  FROM s3.lineitem_1t_part
  UNION ALL SELECT * FROM public.lineitem
  UNION ALL SELECT * FROM apg.lineitem
     with no schema binding;
```

Mostrar el recuento de filas en la vista `lineitem_all` con un predicado para limitar los resultados.

```
SELECT count(*) from lineitem_all WHERE l_quantity = 10;

   count
-----------
123373836
```

Averiguar cuántas ventas de un artículo ha habido en enero de cada año.

```
SELECT extract(year from l_shipdate) as year,
       extract(month from l_shipdate) as month,
       count(*) as orders
FROM lineitem_all
WHERE extract(month from l_shipdate) = 1
AND l_quantity < 2
```

```
GROUP BY 1,2
ORDER BY 1,2;

 year | month | orders
------+-------+---------
 1992 |     1 |  196019
 1993 |     1 | 1582034
 1994 |     1 | 1583181
 1995 |     1 | 1583919
 1996 |     1 | 1583622
 1997 |     1 | 1586541
 1998 |     1 | 1583198
 2016 |     1 |   15542
 2017 |     1 |   15414
 2018 |     1 |   15527
 2019 |     1 |     151
```

## Ejemplo de uso de un nombre compuesto por mayúsculas y minúsculas

Para consultar una base de datos PostgreSQL remota compatible que tenga un nombre de base de datos, esquema, tabla o columna compuesto por mayúsculas y minúsculas, establezca `enable_case_sensitive_identifier` como `true`. Para obtener más información sobre este parámetro de sesión, consulte enable_case_sensitive_identifier.

```
SET enable_case_sensitive_identifier TO TRUE;
```

Por lo general, los nombres de bases de datos y esquemas se escriben en minúsculas. En el siguiente ejemplo, se muestra cómo puede conectarse a una base de datos PostgreSQL remota compatible que tenga nombres escritos en minúsculas para la base de datos y el esquema, por un lado, y nombres compuestos por mayúsculas y minúsculas para la tabla y la columna, por el otro.

Cree un esquema externo que referencie una base de datos de Aurora PostgreSQL con el nombre de la base de datos (`dblower`) y del esquema (`schemalower`) en minúsculas.

```
CREATE EXTERNAL SCHEMA apg_lower
FROM POSTGRES
DATABASE 'dblower' SCHEMA 'schemalower'
URI 'endpoint to aurora hostname'
IAM_ROLE 'arn:aws:iam::123456789012:role/Redshift-SecretsManager-RO'
SECRET_ARN 'arn:aws:secretsmanager:us-west-2:123456789012:secret:federation/test/
dataplane-apg-creds-YbVKQw';
```

En la sesión en la que se ejecuta la consulta, establezca `enable_case_sensitive_identifier` como `true`.

```
SET enable_case_sensitive_identifier TO TRUE;
```

Ejecute una consulta federada para seleccionar todos los datos de la base de datos PostgreSQL. La tabla (`MixedCaseTab`) y la columna (`MixedCaseName`) tienen nombres compuestos por mayúsculas y minúsculas. El resultado es una fila (`Harry`).

```
select * from apg_lower."MixedCaseTab";
```

```
 MixedCaseName
-------
 Harry
```

En el siguiente ejemplo, se muestra cómo puede conectarse a una base de datos PostgreSQL remota compatible que tenga nombres de base de datos, esquema, tabla y columna compuestos por mayúsculas y minúsculas.

Establezca `enable_case_sensitive_identifier` como `true` antes de crear el esquema externo. Si `enable_case_sensitive_identifier` no se establece como `true` antes de crear el esquema externo, se produce un error del tipo que indica que la base de datos no existe.

Cree un esquema externo que referencie una base de datos de Aurora PostgreSQL con el nombre de la base de datos (`UpperDB`) y del esquema (`UpperSchema`) compuesto por mayúsculas y minúsculas.

```
CREATE EXTERNAL SCHEMA apg_upper
FROM POSTGRES
DATABASE 'UpperDB' SCHEMA 'UpperSchema'
URI 'endpoint to aurora hostname'
IAM_ROLE 'arn:aws:iam::123456789012:role/Redshift-SecretsManager-RO'
SECRET_ARN 'arn:aws:secretsmanager:us-west-2:123456789012:secret:federation/test/
dataplane-apg-creds-YbVKQw';
```

Ejecute una consulta federada para seleccionar todos los datos de la base de datos PostgreSQL. La tabla (`MixedCaseTab`) y la columna (`MixedCaseName`) tienen nombres compuestos por mayúsculas y minúsculas. El resultado es una fila (`Harry`).

```
select * from apg_upper."MixedCaseTab";
```

```
 MixedCaseName
-------
 Harry
```

## Ejemplo de uso de una consulta federada con MySQL

En el siguiente ejemplo, se muestra cómo configurar una consulta federada que referencie una base de datos de Aurora MySQL. En este ejemplo se muestra cómo funcionan las consultas federadas. Para ejecutarla en su propio entorno, cámbiela para que se ajuste a su entorno. Para obtener información sobre los requisitos previos para ello, consulte Introducción al uso de consultas federadas en MySQL.

Este ejemplo depende de los siguientes requisitos previos:

- Un secreto que se configuró en Secrets Manager para la base de datos de Aurora MySQL. Se referencia este secreto en las políticas de acceso y los roles de IAM. Para obtener más información, consulte Creación de un secreto y rol de IAM para utilizar consultas federadas.
- Un grupo de seguridad que se configura a partir de la vinculación de Amazon Redshift y Aurora MySQL.

Cree un esquema externo que referencie una base de datos de Aurora MySQL.

```
CREATE EXTERNAL SCHEMA amysql
FROM MYSQL
DATABASE 'functional'
URI 'endpoint to remote hostname'
IAM_ROLE 'arn:aws:iam::123456789012:role/Redshift-SecretsManager-RO'
SECRET_ARN 'arn:aws:secretsmanager:us-west-2:123456789012:secret:federation/test/
dataplane-apg-creds-YbVKQw';
```

Ejecute una expresión SQL SELECT de ejemplo de la tabla de Aurora MySQL para mostrar una fila de la tabla de empleados de Aurora MySQL.

```
SELECT level FROM amysql.employees LIMIT 1;
```

```
 level
-------
     8
```

# Diferencias de los tipos de datos entre Amazon Redshift y las bases de datos PostgreSQL y MySQL compatibles

En la siguiente tabla, se muestra la asignación de un tipo de datos de Amazon Redshift al tipo de datos de Amazon RDS PostgreSQL o Aurora PostgreSQL correspondiente.

| Tipos de datos de Amazon Redshift | Tipo de datos de RDS PostgreSQL o Aurora PostgreSQL | Descripción |
|---|---|---|
| SMALLINT | SMALLINT | Entero firmado de dos bytes |
| INTEGER | INTEGER | Entero firmado de cuatro bytes |
| BIGINT | BIGINT | Entero firmado de ocho bytes |
| DECIMAL | DECIMAL | Numérico exacto de precisión seleccion able |
| REAL | REAL | Número en coma flotante de precisión única |
| DOUBLE PRECISION | DOUBLE PRECISION | Número en coma flotante de precisión doble |
| BOOLEAN | BOOLEAN | Booleano lógico (true/ false) |

| Tipos de datos de Amazon Redshift | Tipo de datos de RDS PostgreSQL o Aurora PostgreSQL | Descripción |
|---|---|---|
| CHAR | CHAR | Cadena de caracteres de longitud fija |
| VARCHAR | VARCHAR | Cadena de caracteres de longitud variable con un límite definido por el usuario |
| FECHA | FECHA | Fecha de calendario (año, mes, día) |
| MARCA DE TIEMPO | MARCA DE TIEMPO | Fecha y hora (sin zona horaria) |
| TIMESTAMPTZ | TIMESTAMPTZ | Fecha y hora (con zona horaria) |
| GEOMETRY | PostGIS GEOMETRY | Datos espaciales |

Los siguientes tipos de datos de RDS PostgreSQL y Aurora PostgreSQL se convierten a VARCHAR(64K) en Amazon Redshift:

- JSON, JSONB
- Matrices
- BIT, BIT VARYING
- BYTEA
- Tipos compuestos
- Tipos de fecha y hora INTERVAL, TIME, TIME WITH TIMEZONE
- Tipos enumerados
- Tipos monetarios
- Tipos de direcciones de red
- Tipos numéricos SERIAL, BIGSERIAL, SMALLSERIAL y MONEY

- Tipos de identificador de objeto
- Tipo pg_lsn
- Pseudotipos
- Tipos de rangos
- Tipos de búsqueda de texto
- TXID_SNAPSHOT
- UUID
- Tipo XML

En la siguiente tabla, se muestra la asignación de un tipo de datos de Amazon Redshift al tipo de datos de Amazon RDS MySQL o Aurora MySQL correspondiente.

| Tipos de datos de Amazon Redshift | Tipo de datos de RDS MySQL o Aurora MySQL | Descripción |
|---|---|---|
| BOOLEAN | TINYINT(1) | Booleano lógico (true o false) |
| SMALLINT | TINYINT(UNSIGNED) | Entero firmado de dos bytes |
| SMALLINT | SMALLINT | Entero firmado de dos bytes |
| INTEGER | SMALLINT UNSIGNED | Entero firmado de cuatro bytes |
| INTEGER | MEDIUMINT (UNSIGNED) | Entero firmado de cuatro bytes |
| INTEGER | INT | Entero firmado de cuatro bytes |
| BIGINT | INT UNSIGNED | Entero firmado de ocho bytes |

| Tipos de datos de Amazon Redshift | Tipo de datos de RDS MySQL o Aurora MySQL | Descripción |
|---|---|---|
| BIGINT | BIGINT | Entero firmado de ocho bytes |
| DECIMAL | BIGINT UNSIGNED | Numérico exacto de precisión seleccion able |
| DECIMAL | DECIMAL(M,D) | Numérico exacto de precisión seleccion able |
| REAL | FLOAT | Número en coma flotante de precisión única |
| DOUBLE PRECISION | DOBLE | Número en coma flotante de precisión doble |
| CHAR | CHAR | Cadena de caracteres de longitud fija |
| VARCHAR | VARCHAR | Cadena de caractere s de longitud variable con un límite definido por el usuario |
| FECHA | FECHA | Fecha de calendario (año, mes, día) |
| HORA | HORA | Hora (sin zona horaria) |
| MARCA DE TIEMPO | MARCA DE TIEMPO | Fecha y hora (sin zona horaria) |

| Tipos de datos de Amazon Redshift | Tipo de datos de RDS MySQL o Aurora MySQL | Descripción |
|---|---|---|
| MARCA DE TIEMPO | DATETIME | Hora (sin zona horaria) |
| VARCHAR(4) | YEAR | Carácter de longitud variable que represent a el año |

Se produce un error cuando los datos TIME están fuera de rango (00:00:00 – 24:00:00).

Los siguientes tipos de datos de RDS MySQL y Aurora MySQL se convierten a VARCHAR(64K) en Amazon Redshift:

- BIT
- BINARIO
- VARBINARY
- TINYBLOB, BLOB, MEDIUMBLOB, LONGBLOB
- TINYTEXT, TEXT, MEDIUMTEXT, LONGTEXT
- ENUM
- SET
- SPATIAL

# Consideraciones al obtener acceso a los datos federados con Amazon Redshift

Algunas características de Amazon Redshift que no admiten el acceso a datos federados. Puede encontrar limitaciones y consideraciones relacionadas a continuación.

A continuación, se indican las limitaciones y las consideraciones que se aplican cuando se utilizan las consultas federadas con Amazon Redshift:

- Las consultas federadas admiten acceso de lectura a orígenes de datos externos. No se pueden escribir ni crear objetos de base de datos en el origen de datos externo.

- En algunos casos, puede obtener acceso a una base de datos del clúster de Amazon RDS o Aurora DB en una región de AWS distinta a la de Amazon Redshift. En estos casos, normalmente se generan cargos por latencia de red y facturación por la transferencia de datos entre regiones de AWS. Se recomienda utilizar una base de datos global de Aurora con un punto de enlace local en la misma región de AWS que su clúster de Amazon Redshift. Las bases de datos globales de Aurora utilizan una infraestructura dedicada para la replicación basada en el almacenamiento en dos regiones de AWS cualesquiera con una latencia típica de menos de 1 segundo.
- Tenga en cuenta el costo de acceso al clúster de Amazon RDS o Aurora DB. Por ejemplo, cuando se utiliza esta característica para acceder al clúster de Aurora DB, los cargos del clúster de Aurora DB se basan en IOPS.
- Las consultas federadas no admiten el acceso a Amazon Redshift desde RDS o el clúster de Aurora DB.
- Las consultas federadas solo están disponibles en las regiones de AWS en las que estén disponibles Amazon Redshift y Amazon RDS o el clúster de Aurora DB.
- Las consultas federadas actualmente no son compatibles con `ALTER SCHEMA`. Para cambiar un esquema, utilice `DROP` y, a continuación, `CREATE EXTERNAL SCHEMA`.
- Las consultas federadas no funcionan con el escalado de simultaneidad.
- Por el momento, las consultas federadas no admiten el acceso a través de un encapsulador de datos externo de PostgreSQL.
- Las consultas federadas a RDS MySQL o Aurora MySQL admiten el aislamiento de transacciones en el nivel READ COMMITTED.
- Si no se especifica, Amazon Redshift se conecta a RDS para MySQL o Aurora MySQL en el puerto 3306. Confirme el número de puerto de MySQL antes de crear un esquema externo para MySQL.
- Si no se especifica, Amazon Redshift se conecta a RDS PostgreSQL o Aurora PostgreSQL en el puerto 5432. Confirme el número de puerto de PostgreSQL antes de crear un esquema externo para PostgreSQL.
- Cuando se obtienen los tipos de datos TIMESTAMP y DATE de MySQL, los valores cero se tratan como NULL.
- Si se utiliza un punto de conexión de lector de base de datos de clúster de Aurora DB, puede producirse un error de “instantánea no válida”. Esto puede evitarse con uno de los siguientes métodos:

- Utilice un punto de conexión de instancia de clúster de Aurora DB específico (en lugar de utilizar el punto de conexión del clúster de Aurora DB). Este método utiliza el aislamiento de transacciones REPEATABLE READ para los resultados de la base de datos PostgreSQL.
- Utilice un punto de conexión de lector de clúster de Aurora DB y establezca `pg_federation_repeatable_read` en false durante la sesión. Este método utiliza el aislamiento de transacciones REPEATABLE READ para los resultados de la base de datos PostgreSQL. Para obtener más información sobre puntos de conexión de lector de clúster de Aurora DB, consulte Tipos de puntos de conexión de clúster de Aurora DB en la Guía del usuario de Amazon Aurora. Para obtener más información sobre `pg_federation_repeatable_read`, consulte pg_federation_repeatable_read.

A continuación, se presentan consideraciones en torno a las transacciones realizadas cuando se usan las consultas federadas a bases de datos PostgreSQL:

- Si una consulta consta de tablas federadas, el nodo principal inicia una transacción READ ONLY REPEATABLE READ en la base de datos remota. Esta transacción permanece todo el tiempo que dura la transacción de Amazon Redshift.
- El nodo principal crea una instantánea de la base de datos remota llamando a `pg_export_snapshot` y realiza un bloqueo de lectura en las tablas afectadas.
- Un nodo de cálculo inicia una transacción y utiliza la instantánea creada en el nodo principal para emitir consultas a la base de datos remota.

## Versiones compatibles de bases de datos federadas

Un esquema externo de Amazon Redshift puede referenciar una base de datos de RDS PostgreSQL o Aurora PostgreSQL externa. En este caso, se aplican estas limitaciones:

- Cuando se crea un esquema externo que consulte el clúster de Aurora DB, la versión de la base de datos de Aurora PostgreSQL debe ser 9.6 o una posterior.
- Cuando se crea un esquema externo que consulte Amazon RDS, la versión de la base de datos de Amazon RDS PostgreSQL debe ser 9.6 o una posterior.

Un esquema externo de Amazon Redshift puede referenciar una base de datos de RDS MySQL o Aurora MySQL externa. En este caso, se aplican estas limitaciones:

- Cuando se crea un esquema externo que consulte el clúster de Aurora DB, la versión de la base de datos de Aurora MySQL debe ser 5.6 o una posterior.
- Cuando se cree un esquema externo que revise Amazon RDS, la versión de la base de datos de RDS MySQL debe ser 5.6 o una posterior.

# Amazon Redshift Spectrum

En esta sección se describe cómo utilizar Redshift Spectrum para leer datos de Amazon S3 de forma eficiente.

Con Amazon Redshift Spectrum, puede consultar y recuperar de manera eficiente los datos estructurados y semiestructurados de los archivos en Amazon S3 sin tener que cargar los datos en tablas de Amazon Redshift. Las consultas de Redshift Spectrum emplean un paralelismo a gran escala para ejecutar muy rápido grandes conjuntos de datos. Gran parte del procesamiento se produce en la capa de Redshift Spectrum, mientras que la mayoría de los datos permanece en Amazon S3. Varios clústeres pueden consultar el mismo conjunto de datos en Amazon S3 al mismo tiempo sin tener que generar copias de los datos para cada clúster.

Temas


- Información general sobre Amazon Redshift Spectrum
- Introducción a Amazon Redshift Spectrum
- Políticas de IAM para Amazon Redshift Spectrum
- Redshift Spectrum y AWS Lake Formation
- Archivos de datos para consultas en Amazon Redshift Spectrum
- Esquemas externos en Amazon Redshift Spectrum
- Tablas externas para Redshift Spectrum
- Uso de tablas de Apache Iceberg con Amazon Redshift
- Rendimiento de las consultas de Amazon Redshift Spectrum
- Opciones de gestión de datos
- Ejemplo: Realización de subconsultas correlacionadas en Redshift Spectrum
- Métricas en Amazon Redshift Spectrum
- Solución de problemas de consultas en Amazon Redshift Spectrum
- Tutorial: Consulta de datos anidados con Amazon Redshift Spectrum


## Información general sobre Amazon Redshift Spectrum

En este tema se describe cómo utilizar Redshift Spectrum para leer de Amazon S3 de forma eficiente.

Amazon Redshift Spectrum se encuentra en servidores de Amazon Redshift dedicados que no dependen del clúster. Amazon Redshift inserta muchas tareas de computación intensiva, como el filtrado y la agregación de predicados, en la capa de Redshift Spectrum. Por lo tanto, las consultas de Redshift Spectrum utilizan mucha menos capacidad de procesamiento de su clúster que otras consultas. Redshift Spectrum también escala de manera inteligente. Según la demanda de las consultas, Redshift Spectrum puede utilizar miles de instancias para aprovechar el procesamiento masivo en paralelo.

Puede crear tablas de Redshift Spectrum al definir la estructura para sus archivos y registrarlos como tablas en un catálogo de datos externo. El catálogo de datos externo puede ser AWS Glue, el catálogo de datos que incluye Amazon Athena o su propio metastore de Apache Hive. Puede crear y administrar tablas externas desde Amazon Redshift mediante comandos de lenguaje de definición de datos (DDL) o cualquier otra herramienta que se conecte al catálogo de datos externo. Los cambios en el catálogo de datos externo están disponibles de inmediato en cualquiera de los clústeres de Amazon Redshift.

Opcionalmente, puede particionar las tablas externas en una o más columnas. Definir las particiones como parte de la tabla externa puede mejorar el rendimiento. La mejora se produce porque el optimizador de consultas de Amazon Redshift elimina las particiones que no tienen datos para la consulta.

Las vistas materializadas en las tablas de Spectrum pueden mejorar considerablemente el costo y el rendimiento. Para obtener más información, consulte Vistas materializadas de tablas de lagos de datos externos en Amazon Redshift Spectrum.

Después de que se hayan definido las tablas de Redshift Spectrum, puede consultar y combinar las tablas de la misma manera en que lo hace con cualquier otra tabla de Amazon Redshift. Redshift Spectrum no permite realizar operaciones de actualización en tablas externas. Puede agregar tablas de Redshift Spectrum a distintos clústeres de Amazon Redshift y consultar los mismos datos de Amazon S3 desde cualquier clúster que esté en la misma región de AWS. Cuando actualiza los archivos de datos de Amazon S3, los datos están disponibles de inmediato para consultarlos desde cualquiera de los clústeres de Amazon Redshift.

El catálogo de datos de AWS Glue al que obtiene acceso podría estar cifrado para aumentar la seguridad. Si el catálogo de AWS Glue está cifrado, necesitará la clave de AWS Key Management Service (AWS KMS) para que AWS Glue pueda obtener acceso al catálogo de AWS Glue. El cifrado del catálogo de AWS Glue no está disponible en todas las regiones de AWS. Para obtener una lista de las regiones de AWS admitidas, consulte Cifrado y acceso seguro para AWS Glue en la Guía para desarrolladores de AWS Glue. Para obtener más información sobre el cifrado del catálogo

de datos de AWS Glue, consulte Cifrado de su catálogo de datos de AWS Glue en la Guía para desarrolladores de AWS Glue.

> **Note**
>
> No puede ver los detalles de las tablas de Redshift Spectrum con los mismos recursos que utiliza para las tablas de Amazon Redshift estándar, como PG_TABLE_DEF, STV_TBL_PERM, PG_CLASS o information_schema. Si su herramienta de análisis o inteligencia empresarial no reconoce las tablas externas de Redshift Spectrum, configure la aplicación para consultar SVV_EXTERNAL_TABLES y SVV_EXTERNAL_COLUMNS.

## Regiones de Amazon Redshift Spectrum

Redshift Spectrum está disponible en las Regiones de AWS donde esté disponible Amazon Redshift, a menos que se indique lo contrario en la documentación específica de las regiones. Para conocer la disponibilidad de Región de AWS en regiones comerciales, consulte Puntos de conexión de servicio para la API de Redshift en la Referencia general de Amazon Web Services.

## Limitaciones de Amazon Redshift Spectrum

En este tema se describen las limitaciones para usar Redshift Spectrum.

Tenga en cuenta las siguientes consideraciones cuando utilice Redshift Spectrum:

- El clúster de Amazon Redshift y el bucket de Amazon S3 deben estar en la misma región de AWS.
- Redshift Spectrum no admite el enrutamiento de VPC mejorado con clústeres aprovisionados. Para acceder a sus datos de Amazon S3, es posible que deba realizar pasos de configuración adicionales. Para obtener más información, consulte Redshift Spectrum y enrutamiento de VPC mejorado en la Guía de administración de Amazon Redshift.
- Redshift Spectrum es compatible con los alias de puntos de acceso de Amazon S3. Para obtener más información, consulte Uso de un alias de estilo bucket en su punto de acceso en la Guía del usuario de Amazon Simple Storage Service. No obstante, Redshift Spectrum no admite la VPC con alias de punto de acceso de Amazon S3. Para obtener más información, consulte Redshift Spectrum y enrutamiento de VPC mejorado en la Guía de administración de Amazon Redshift.
- No puede realizar operaciones de actualización o eliminación en tablas externas. Para crear una nueva tabla externa en el esquema especificado, puede utilizar CREATE EXTERNAL TABLE. Para obtener más información acerca de CREATE EXTERNAL TABLE AS, consulte CREATE

EXTERNAL TABLE. Para insertar los resultados de una consulta SELECT en tablas externas existentes en catálogos externos, puede utilizar INSERT (tabla externa). Para obtener más información acerca de INSERT (tabla externa), consulte INSERT (tabla externa).

- A menos que esté utilizando un AWS Glue Data Catalog habilitado para AWS Lake Formation, no puede controlar los permisos de usuario en una tabla externa. En cambio, sí puede conceder y revocar permisos en el esquema externo. Para obtener más información sobre el uso de AWS Lake Formation, consulte Redshift Spectrum y AWS Lake Formation.
- Para ejecutar consultas de Redshift Spectrum, el usuario de la base de datos debe tener permiso para crear tablas temporales en ella. En el siguiente ejemplo, se concede un permiso temporal para la base de datos `spectrumdb` al grupo de usuarios `spectrumusers`.

```
grant temp on database spectrumdb to group spectrumusers;
```

  Para obtener más información, consulte GRANT.
- Cuando utilice el catálogo de datos de Athena o de AWS Glue como almacén de metadatos, consulte Cuotas y límites en la Guía de administración de Amazon Redshift.
- Redshift Spectrum no es compatible con Amazon EMR con Kerberos.

# Introducción a Amazon Redshift Spectrum

En este tutorial, aprenderá a utilizar Amazon Redshift Spectrum para consultar datos directamente en los archivos de Amazon S3. Si ya tiene un clúster y un cliente SQL, puede completar este tutorial con la configuración mínima.

> **Note**
>
> Las consultas de Redshift Spectrum ocasionan cargos adicionales. El costo de ejecución de las consultas de muestra en este tutorial es nominal. Para obtener más información acerca de los precios, consulte Precios de Amazon Redshift Spectrum.

## Requisitos previos

Para usar Redshift Spectrum, necesita un clúster de Amazon Redshift y un cliente SQL que se conecte al clúster, de modo que pueda ejecutar comandos SQL. El clúster y los archivos de datos de Amazon S3 deben estar en la misma Región de AWS.

Para obtener más información sobre cómo crear un clúster de Amazon Redshift, consulte Introducción a los almacenamientos de datos aprovisionados de Amazon Redshift en la Guía de introducción a Amazon Redshift. Para obtener información sobre las formas de conectarse a un clúster, consulte Conexión a los almacenamientos de datos de Amazon Redshift en la Guía de introducción a Amazon Redshift.

En algunos de los ejemplos siguientes, los datos de muestra se encuentran en la región Este de EE. UU. (Norte de Virginia) (`us-east-1`), por lo que necesita un clúster que también esté en `us-east-1`. O bien, puede usar Amazon S3 para copiar objetos de datos desde los siguientes buckets y carpetas en su bucket en el Región de AWS donde se encuentra el clúster:

- `s3://redshift-downloads/tickit/spectrum/customers/*`
- `s3://redshift-downloads/tickit/spectrum/sales_partition/*`
- `s3://redshift-downloads/tickit/spectrum/sales/*`
- `s3://redshift-downloads/tickit/spectrum/salesevent/*`

Ejecute un comando de Amazon S3 similar al siguiente para copiar los datos de muestra que se encuentran en el Este de EE. UU. (Norte de Virginia) en su Región de AWS. Antes de ejecutar el comando, cree el bucket y las carpetas del bucket para que coincidan con el comando copy de Amazon S3. El resultado del comando copy de Amazon S3 confirma que los archivos se han copiado en el *`nombre-bucket`* de la Región de AWS que desea.

```
aws s3 cp s3://redshift-downloads/tickit/spectrum/ s3://bucket-name/tickit/spectrum/ --
copy-props none --recursive
```

## Introducción a Redshift Spectrum con AWS CloudFormation

Como alternativa a los siguientes pasos, puede acceder a la plantilla DataLake de AWS CloudFormation de Redshift Spectrum para crear una pila con un bucket de Amazon S3 en la que luego pueda realizar consultas. Para obtener más información, consulte Lanzamiento de su pila de AWS CloudFormation y, a continuación, consulta de sus datos en Amazon S3.

## Introducción paso a paso a Redshift Spectrum

Para empezar a usar Amazon Redshift Spectrum, siga los pasos que se indican a continuación:

- Paso 1. Crear un rol de IAM para Amazon Redshift
- Paso 2: Asociar el rol de IAM a su clúster

- Paso 3: Crear un esquema externo y una tabla externa
- Paso 4: Consultar sus datos en Amazon S3

## Paso 1. Crear un rol de IAM para Amazon Redshift

El clúster necesita autorización para obtener acceso al catálogo de datos externo de AWS Glue o Amazon Athena, y a los archivos de datos de Amazon S3. Para proporcionar esa autorización, se referencia un rol de AWS Identity and Access Management (IAM) que se adjunta a su clúster. Para obtener más información acerca del uso de roles con Amazon Redshift, consulte Autorización de uso de las operaciones COPY y UNLOAD mediante roles de IAM.

> **Note**
>
> En ciertos casos, puede migrar su catálogo de datos de Athena a un catálogo de datos de AWS Glue. Puede hacer esto si su clúster está en una región de AWS donde se admite AWS Glue y tiene tablas externas de Redshift Spectrum en el catálogo de datos de Athena. Para usar el catálogo de datos de AWS Glue con Redshift Spectrum, es posible que tenga que cambiar las políticas de IAM. Para obtener más información, consulte Actualización al catálogo de datos de AWS Glue en la Guía del usuario de Athena.

Cuando cree un rol para Amazon Redshift, elija uno de los siguientes enfoques:

- Si está utilizando Redshift Spectrum con un catálogo de datos de Athena o un catálogo de datos de AWS Glue, siga los pasos descritos en Para crear un rol de IAM para Amazon Redshift.
- Si está utilizando Redshift Spectrum con un AWS Glue Data Catalog que está habilitado para AWS Lake Formation, siga los pasos descritos en estos procedimientos:
  - Para crear un rol de IAM para Amazon Redshift mediante un AWS Glue Data Catalog habilitado para AWS Lake Formation
  - Para otorgar permisos SELECT en la tabla para realizar consultas en la base de datos de Lake Formation

Para crear un rol de IAM para Amazon Redshift

1. Abra la consola de IAM.
2. Seleccione Roles en el panel de navegación.

3. Elija Create role (Crear rol).
4. Elija Servicio de AWS como entidad de confianza y, a continuación, elija Redshift como caso de uso.
5. En Caso de uso para otros Servicios de AWS, elija Redshift - Personalizable y, a continuación, elija Siguiente.
6. Aparece la página Add permissions policy (Agregar política de permisos). Elija `AmazonS3ReadOnlyAccess` y `AWSGlueConsoleFullAccess`, si está utilizando el catálogo de datos AWS Glue. O bien, elija `AmazonAthenaFullAccess` si está utilizando el catálogo de datos de Athena. Elija Next (Siguiente).

   > **Note**
   >
   > La política `AmazonS3ReadOnlyAccess` proporciona al clúster acceso de solo lectura a todos los buckets de Amazon S3. Para conceder acceso solo al bucket de datos de muestra de AWS, cree una política nueva y agregue los siguientes permisos.

   ```
   {
       "Version": "2012-10-17",
       "Statement": [
           {
               "Effect": "Allow",
               "Action": [
                   "s3:Get*",
                   "s3:List*"
               ],
               "Resource": "arn:aws:s3:::redshift-downloads/*"
           }
       ]
   }
   ```

7. En Role Name (Nombre de la función), escriba un nombre para la función; por ejemplo, **`myspectrum_role`**.
8. Revise la información y seleccione Create role (Crear función).
9. Seleccione Roles en el panel de navegación. Elija el nombre de la nueva función para ver el resumen y copie el ARN de función en el portapapeles. Esta valor es el nombre de recurso de Amazon (ARN) de la función que acaba de crear. Usa ese valor cuando crea tablas externas para referenciar sus archivos de datos en Amazon S3.

Para crear un rol de IAM para Amazon Redshift mediante un AWS Glue Data Catalog habilitado para AWS Lake Formation

1. Abra la consola de IAM en https://console.aws.amazon.com/iam/.
2. En el panel de navegación, seleccione Policies (Políticas).

   Si es la primera vez que elige Policies (Políticas), aparecerá la página Welcome to Managed Policies (Bienvenido a políticas administradas). Elija Get Started (Comenzar).
3. Elija Create Policy (Crear política).
4. Elija crear la política en la pestaña JSON.
5. Pegue el siguiente documento de política JSON, el cual concede acceso al catálogo de datos, pero deniega el permiso de administrador para Lake Formation.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "RedshiftPolicyForLF",
            "Effect": "Allow",
            "Action": [
                "glue:*",
                "lakeformation:GetDataAccess"
            ],
            "Resource": "*"
        }
    ]
}
```

6. Cuando haya terminado, seleccione Review (Revisar) para revisar la política. El validador de políticas notifica los errores de sintaxis.
7. En la página Review policy (Revisar política), en el campo Name (Nombre), ingrese **`myspectrum_policy`** para asignarle un nombre a la política que está creando. (Opcional) Introduzca una descripción. Revise el Summary (Resumen) de la política para ver los permisos concedidos por su política. A continuación, elija Create policy (Crear política) para guardar su trabajo.

   Después de crear una política, puede proporcionar acceso a los clientes.

Para dar acceso, agregue permisos a los usuarios, grupos o roles:

- Usuarios y grupos en AWS IAM Identity Center:

  Cree un conjunto de permisos. Siga las instrucciones de Creación de un conjunto de permisos en la Guía del usuario de AWS IAM Identity Center.

- Usuarios gestionados en IAM a través de un proveedor de identidades:

  Cree un rol para la federación de identidades. Siga las instrucciones descritas en Creación de un rol para un proveedor de identidad de terceros (federación) en la Guía del usuario de IAM.

- Usuarios de IAM:
  - Cree un rol que el usuario pueda aceptar. Siga las instrucciones descritas en Creación de un rol para un usuario de IAM en la Guía del usuario de IAM.
  - (No recomendado) Adjunte una política directamente a un usuario o añada un usuario a un grupo de usuarios. Siga las instrucciones descritas en Adición de permisos a un usuario (consola) de la Guía del usuario de IAM.

Para otorgar permisos SELECT en la tabla para realizar consultas en la base de datos de Lake Formation

1. Abra la consola de Lake Formation en https://console.aws.amazon.com/lakeformation/.
2. En el panel de navegación, elija Permisos de lago de datos y, a continuación, elija Otorgar.
3. Siga las instrucciones de Otorgar permisos de tabla mediante el método de recurso designado de la Guía para desarrolladores de AWS Lake Formation. Proporcione la información siguiente:
   - En IAM role (Rol de IAM), elija el rol de IAM que creó, `myspectrum_role`. Cuando se ejecuta el editor de consultas de Amazon Redshift, este utiliza el rol de IAM para dar permiso a los datos.

     > **Note**
     >
     > Si desea otorgar permisos SELECT en la tabla de un catálogo de datos habilitado para Lake Formation para realizar consultas, haga lo siguiente:
     >
     > - Registre la ruta para los datos en Lake Formation.
     > - Otorgue a los usuarios permisos para la ruta en Lake Formation.
     > - Las tablas creadas se pueden encontrar en la ruta registrada en Lake Formation.

4. Elija Conceder.

> **Important**
>
> Como práctica recomendada, permita el acceso únicamente a los objetos de Amazon S3 subyacentes a través de permisos de Lake Formation. Para evitar accesos no aprobados, quite cualquier permiso que haya otorgado a los objetos de Amazon S3 fuera de Lake Formation. Si ya ha accedido a objetos de Amazon S3 antes de configurar Lake Formation, quite las políticas de IAM o los permisos de los buckets que se hayan configurado con anterioridad. Para obtener más información, consulte Actualización de permisos de datos de AWS Glue en el modelo de AWS Lake Formation y Permisos de Lake Formation.

## Paso 2: Asociar el rol de IAM a su clúster

Ahora, dispone de un rol de IAM que autoriza a Amazon Redshift a acceder al catálogo de datos externo y a Amazon S3 por usted. A esta altura, debe asociar ese rol a su clúster de Amazon Redshift.

Asociar un rol de IAM con un clúster

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. En el menú de navegación, elija Clusters (Clústeres) y, a continuación, elija el nombre del clúster que desea actualizar.
3. En Actions (Acciones), seleccione Manage IAM roles (Administrar funciones de IAM). Aparece la página IAM roles (Roles de IAM).
4. Seleccione Enter ARN (Introducir ARN) e introduzca un ARN o un rol de IAM o seleccione un rol de IAM de la lista. Luego elija Add IAM role (Agregar un rol de IAM) para agregarlo a la lista de Attached IAM roles (Roles de IAM agregados).
5. Seleccione Done (Hecho) para asociar el rol de IAM al clúster. El clúster se modifica para completar el cambio.

## Paso 3: Crear un esquema externo y una tabla externa

Crear tablas externas en un esquema externo. El esquema externo referencia una base de datos en el catálogo de datos externo y proporciona el ARN del rol de IAM que autoriza a su clúster a obtener

acceso a Amazon S3 en su nombre. Puede crear una base de datos externa en un catálogo de datos de Amazon Athena, AWS Glue Data Catalog, o un metastore de Apache Hive, como Amazon EMR. En este ejemplo, crea la base de datos externa en un catálogo de datos de Amazon Athena cuando crea el esquema externo de Amazon Redshift. Para obtener más información, consulte Esquemas externos en Amazon Redshift Spectrum.

Pasos para crear un esquema externo y una tabla externa

1. Para crear un esquema externo, reemplace el ARN de la función de IAM en el siguiente comando por el ARN de la función que creó en el Paso 1. Luego, ejecute el comando en cliente SQL.

```
create external schema myspectrum_schema
from data catalog
database 'myspectrum_db'
iam_role 'arn:aws:iam::123456789012:role/myspectrum_role'
create external database if not exists;
```

2. Para crear una tabla externa, ejecute el siguiente comando CREATE EXTERNAL TABLE.

> **Note**
>
> Su clúster y el bucket de Amazon S3 deben estar en el mismo Región de AWS. Para este ejemplo de comando CREATE EXTERNAL TABLE, el bucket de Amazon S3 con los datos de muestra se encuentra en la Región de AWS Este de EE. UU. (Norte de Virginia). Para ver los datos de origen, descargue el archivo `sales_ts.000`.
> Puede modificar este ejemplo para que se ejecute en otra Región de AWS. Cree un bucket de Amazon S3 en la Región de AWS que desee. Cree los datos de ventas con un comando copy de Amazon S3. A continuación, actualice la opción de ubicación en el comando `CREATE EXTERNAL TABLE` de ejemplo para su bucket.
>
> ```
> aws s3 cp s3://redshift-downloads/tickit/spectrum/sales/ s3://bucket-name/
> tickit/spectrum/sales/ --copy-props none --recursive
> ```
>
> El resultado del comando copy de Amazon S3 confirma que el archivo se ha copiado en el *nombre-bucket* de la Región de AWS que desea.

```
copy: s3://redshift-downloads/tickit/spectrum/sales/sales_ts.000 to
 s3://bucket-name/tickit/spectrum/sales/sales_ts.000
```

```
create external table myspectrum_schema.sales(
salesid integer,
listid integer,
sellerid integer,
buyerid integer,
eventid integer,
dateid smallint,
qtysold smallint,
pricepaid decimal(8,2),
commission decimal(8,2),
saletime timestamp)
row format delimited
fields terminated by '\t'
stored as textfile
location 's3://redshift-downloads/tickit/spectrum/sales/'
table properties ('numRows'='172000');
```

## Paso 4: Consultar sus datos en Amazon S3

Una vez creadas las tablas externas, puede consultarlas con las mismas instrucciones SELECT que usa para consultar otras tablas de Amazon Redshift. Estas consultas de la instrucción SELECT incluyen tablas combinadas, datos de agregación y filtrado de predicados.

Para consultar los datos en Amazon S3

1. Obtenga la cantidad de filas en la tabla MYSPECTRUM_SCHEMA.SALES.

```
select count(*) from myspectrum_schema.sales;
```

```
count
------
172462
```

2. Como práctica recomendada, es conveniente que guarde las tablas de hechos más grandes en Amazon S3 y las tablas de dimensiones más pequeñas en Amazon Redshift. Si cargó los datos de muestra en Cargar datos, tendrá una tabla llamada EVENT en la base de datos. Si no es así, cree la tabla EVENT utilizando el siguiente comando.

```
create table event(
eventid integer not null distkey,
venueid smallint not null,
catid smallint not null,
dateid smallint not null sortkey,
eventname varchar(200),
starttime timestamp);
```

3. Cargue la tabla EVENT reemplazando el ARN del rol de IAM en el siguiente comando COPY por el ARN del rol que creó en Paso 1. Crear un rol de IAM para Amazon Redshift. Si lo desea, puede descargar y ver los datos de origen para `allevents_pipe.txt`desde un bucket de Amazon S3 en la Región de AWS `us-east-1`.

```
copy event from 's3://redshift-downloads/tickit/allevents_pipe.txt'
iam_role 'arn:aws:iam::123456789012:role/myspectrum_role'
delimiter '|' timeformat 'YYYY-MM-DD HH:MI:SS' region 'us-east-1';
```

En el siguiente ejemplo, se combina la tabla externa de Amazon S3 MYSPECTRUM_SCHEMA.SALES con la tabla local de Amazon Redshift EVENT para obtener el total de ventas de los 10 principales eventos.

```
select top 10 myspectrum_schema.sales.eventid,
 sum(myspectrum_schema.sales.pricepaid) from myspectrum_schema.sales, event
where myspectrum_schema.sales.eventid = event.eventid
and myspectrum_schema.sales.pricepaid > 30
group by myspectrum_schema.sales.eventid
order by 2 desc;
```

```
eventid | sum
--------+---------
    289 | 51846.00
   7895 | 51049.00
   1602 | 50301.00
    851 | 49956.00
   7315 | 49823.00
```

```
   6471 | 47997.00
   2118 | 47863.00
    984 | 46780.00
   7851 | 46661.00
   5638 | 46280.00
```

4. Vea el plan de consulta de la consulta anterior. Observe los pasos `S3 Seq Scan`, `S3 HashAggregate` y `S3 Query Scan` que se ejecutaron sobre los datos de Amazon S3.

```
explain
select top 10 myspectrum_schema.sales.eventid,
 sum(myspectrum_schema.sales.pricepaid)
from myspectrum_schema.sales, event
where myspectrum_schema.sales.eventid = event.eventid
and myspectrum_schema.sales.pricepaid > 30
group by myspectrum_schema.sales.eventid
order by 2 desc;
```

```
QUERY PLAN


------------------------------------------------------------------------------
XN Limit  (cost=1001055770628.63..1001055770628.65 rows=10 width=31)


  ->  XN Merge  (cost=1001055770628.63..1001055770629.13 rows=200 width=31)


        Merge Key: sum(sales.derived_col2)


        ->  XN Network  (cost=1001055770628.63..1001055770629.13 rows=200 width=31)


              Send to leader


              ->  XN Sort  (cost=1001055770628.63..1001055770629.13 rows=200
 width=31)
```

```
                  Sort Key: sum(sales.derived_col2)


                  ->  XN HashAggregate  (cost=1055770620.49..1055770620.99
rows=200 width=31)

                        ->  XN Hash Join DS_BCAST_INNER
(cost=3119.97..1055769620.49 rows=200000 width=31)

                              Hash Cond: ("outer".derived_col1 = "inner".eventid)


                              ->  XN S3 Query Scan sales  (cost=3010.00..5010.50
rows=200000 width=31)

                                    ->  S3 HashAggregate  (cost=3010.00..3010.50
rows=200000 width=16)

                                          ->  S3 Seq Scan myspectrum_schema.sales
location:"s3://redshift-downloads/tickit/spectrum/sales" format:TEXT
(cost=0.00..2150.00 rows=172000 width=16)
                                                Filter: (pricepaid > 30.00)


                              ->  XN Hash  (cost=87.98..87.98 rows=8798 width=4)


                                    ->  XN Seq Scan on event  (cost=0.00..87.98
rows=8798 width=4)
```

## Lanzamiento de su pila de AWS CloudFormation y, a continuación, consulta de sus datos en Amazon S3

Una vez que haya creado un clúster de Amazon Redshift y se haya conectado a él, podrá instalar su plantilla DataLake de AWS CloudFormation de Redshift Spectrum y, a continuación, consultar sus datos.

CloudFormation instala la plantilla DataLake de Introducción a Redshift Spectrum y crea una pila que comprende lo siguiente:

- un rol denominado `myspectrum_role` asociado a su clúster de Redshift

- un esquema externo denominado `myspectrum_schema`
- una tabla externa denominada `sales` en un bucket de Amazon S3
- una tabla de Redshift denominada `event` con datos cargados

Para lanzar la pila DataLake de CloudFormation de Introducción a Redshift Spectrum.

1. Elija Launch CFN stack (Lanzar pila de CFN). La consola de CloudFormation se abre con la plantilla Datalake.yml seleccionada.

   También puede descargar y personalizar la plantilla CFN de introducción a DataLake CloudFormation de Redshift Spectrum y, después, abrir la consola de CloudFormation (https://console.aws.amazon.com/cloudformation) y crear una pila con la plantilla personalizada.
2. Elija Siguiente.
3. En la opción Parameters (Parámetros), ingrese el nombre del clúster de Amazon Redshift, el nombre de la base de datos y su correspondiente nombre de usuario.
4. Elija Next (Siguiente).

   Aparecerán las opciones de la pila.
5. Elija Next (Siguiente) para aceptar la configuración predeterminada.
6. Revise la información y, en la opción Capabilities (Capacidades), elija I acknowledge that AWS CloudFormation might create IAM resources (Acepto que CFN podría crear recursos de IAM).
7. Seleccione Creación de pila.

Si se produce un error durante la creación de la pila, consulte la siguiente información:

- Consulte la pestaña Events (Eventos) de CloudFormation para obtener información que pueda serle de utilidad para resolver el error.
- Elimine la pila DataLake de CloudFormation antes de reintentar la operación.
- Asegúrese de que se ha conectado a su base de datos de Amazon Redshift.
- Asegúrese de que ingresó la información adecuada relativa al nombre del clúster de Amazon Redshift, el nombre de la base de datos y su correspondiente nombre de usuario.

## Consulta de los datos en Amazon S3

Las tablas externas se consultan con las mismas instrucciones SELECT que se utilizan para hacerlo con otras tablas de Amazon Redshift. Estas consultas de la instrucción SELECT incluyen tablas combinadas, datos de agregación y filtrado de predicados.

La siguiente consulta devuelve la cantidad de filas de la tabla externa `myspectrum_schema.sales`.

```
select count(*) from myspectrum_schema.sales;
```

```
count
------
172462
```

## Combinación de una tabla externa con una tabla local

En el siguiente ejemplo, se combina la tabla externa `myspectrum_schema.sales` con la tabla local `event` para obtener el total de ventas de los 10 eventos principales.

```
select top 10 myspectrum_schema.sales.eventid, sum(myspectrum_schema.sales.pricepaid)
 from myspectrum_schema.sales, event
where myspectrum_schema.sales.eventid = event.eventid
and myspectrum_schema.sales.pricepaid > 30
group by myspectrum_schema.sales.eventid
order by 2 desc;
```

```
eventid | sum
--------+---------
    289 | 51846.00
   7895 | 51049.00
   1602 | 50301.00
    851 | 49956.00
   7315 | 49823.00
   6471 | 47997.00
   2118 | 47863.00
    984 | 46780.00
   7851 | 46661.00
   5638 | 46280.00
```

## Visualización del plan de consulta

Vea el plan de consulta de la consulta anterior. Tenga en cuenta los pasos `S3 Seq Scan`, `S3 HashAggregate` y `S3 Query Scan` que se ejecutaron sobre los datos de Amazon S3.

```
explain
select top 10 myspectrum_schema.sales.eventid, sum(myspectrum_schema.sales.pricepaid)
from myspectrum_schema.sales, event
where myspectrum_schema.sales.eventid = event.eventid
and myspectrum_schema.sales.pricepaid > 30
group by myspectrum_schema.sales.eventid
order by 2 desc;
```

```
QUERY PLAN


------------------------------------------------------------------------------
XN Limit  (cost=1001055770628.63..1001055770628.65 rows=10 width=31)


  ->  XN Merge  (cost=1001055770628.63..1001055770629.13 rows=200 width=31)


        Merge Key: sum(sales.derived_col2)


        ->  XN Network  (cost=1001055770628.63..1001055770629.13 rows=200 width=31)


              Send to leader


              ->  XN Sort  (cost=1001055770628.63..1001055770629.13 rows=200 width=31)


                    Sort Key: sum(sales.derived_col2)


                    ->  XN HashAggregate  (cost=1055770620.49..1055770620.99 rows=200
 width=31)
```

```
                          ->  XN Hash Join DS_BCAST_INNER  (cost=3119.97..1055769620.49
rows=200000 width=31)

                                Hash Cond: ("outer".derived_col1 = "inner".eventid)


                                ->  XN S3 Query Scan sales  (cost=3010.00..5010.50
rows=200000 width=31)

                                      ->  S3 HashAggregate  (cost=3010.00..3010.50
rows=200000 width=16)

                                            ->  S3 Seq Scan spectrum.sales
location:"s3://redshift-downloads/tickit/spectrum/sales" format:TEXT
(cost=0.00..2150.00 rows=172000 width=16)
                                                  Filter: (pricepaid > 30.00)


                                ->  XN Hash  (cost=87.98..87.98 rows=8798 width=4)


                                      ->  XN Seq Scan on event  (cost=0.00..87.98
rows=8798 width=4)
```

# Políticas de IAM para Amazon Redshift Spectrum

En este tema se describen los permisos de IAM necesarios para usar Redshift Spectrum.

De manera predeterminada, Amazon Redshift Spectrum utiliza el AWS Glue Data Catalog en las regiones de AWS que admiten AWS Glue. En otras regiones de AWS, Redshift Spectrum utiliza el catálogo de datos de Athena. El clúster necesita autorización para acceder al catálogo de datos externo de AWS Glue o Athena, y a los archivos de datos de Amazon S3. Puede proporcionar esa autorización referenciando un rol de AWS Identity and Access Management (IAM) que esté asociado a su clúster. Si utiliza un metastore de Apache Hive para administrar su catálogo de datos, no necesita proporcionar acceso a Athena.

Puede encadenar roles para que el clúster pueda asumir otros roles no asociados al clúster. Para obtener más información, consulte Encadenamiento de roles de IAM en Amazon Redshift Spectrum.

El catálogo de datos de AWS Glue al que obtiene acceso podría estar cifrado para aumentar la seguridad. Si el catálogo de AWS Glue está cifrado, necesita la clave de AWS KMS para que AWS

Glue acceda al catálogo de AWS Glue. Para obtener más información, consulte Cifrado de su catálogo de datos de AWS Glue en la Guía para desarrolladores de AWS Glue.

Temas

- Permisos de Amazon S3
- Permisos entre cuentas de Amazon S3
- Políticas para conceder o restringir el acceso a través de Redshift Spectrum
- Políticas para conceder los permisos mínimos
- Encadenamiento de roles de IAM en Amazon Redshift Spectrum
- Control del acceso al catálogo de datos de AWS Glue

## Permisos de Amazon S3

Como mínimo, el clúster necesita obtener acceso de tipo GET y LIST a su bucket de Amazon S3. Si el bucket no está en la misma cuenta de AWS que el clúster, el bucket también debe conceder permiso al clúster para acceder a los datos. Para obtener más información, consulte Autorización a Amazon Redshift para obtener acceso a otros servicios de AWS en su nombre.

> **Note**
>
> El bucket de Amazon S3 no puede usar una política de bucket que restrinja el acceso solo desde puntos de enlace de la VPC específicos.

La siguiente política concede acceso de tipo GET y LIST a cualquier bucket de Amazon S3. La política otorga acceso a los buckets de Amazon S3 para Redshift Spectrum, además de operaciones COPY.

```
{
 "Version": "2012-10-17",
 "Statement": [{
  "Effect": "Allow",
  "Action": ["s3:Get*", "s3:List*"],
  "Resource": "*"
 }]
}
```

La siguiente política concede acceso de tipo GET y LIST a un bucket de Amazon S3 denominado `amzn-s3-demo-bucket`.

```
{
 "Version": "2012-10-17",
 "Statement": [{
  "Effect": "Allow",
  "Action": ["s3:Get*", "s3:List*"],
  "Resource": "arn:aws:s3:::amzn-s3-demo-bucket/*"
 }]
}
```

## Permisos entre cuentas de Amazon S3

Para conceder a Redshift Spectrum permiso para acceder a los datos de un bucket de Amazon S3 que pertenece a otra cuenta de AWS, agregue la siguiente política al bucket de Amazon S3. Para obtener más información, consulte Concesión de permisos de bucket entre cuentas.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "Example permissions",
            "Effect": "Allow",
            "Principal": {
                "AWS": "arn:aws:iam::redshift-account:role/spectrumrole"
            },
            "Action": [
                "s3:GetBucketLocation",
                "s3:GetObject",
                "s3:ListMultipartUploadParts",
                "s3:ListBucket",
                "s3:ListBucketMultipartUploads"
            ],
            "Resource": [
                "arn:aws:s3:::bucketname",
                "arn:aws:s3:::bucketname/*"
            ]
        }
    ]
}
```

## Políticas para conceder o restringir el acceso a través de Redshift Spectrum

La siguiente política de ejemplo permite el acceso a los buckets de Amazon S3 para Amazon Redshift. Al agregar esta política a un rol de IAM, recomendamos utilizar un rol de IAM que solo se use para Amazon Redshift. También recomendamos que no reutilice roles existentes ni agregue permisos adicionales a este rol de IAM.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "s3:GetObject",
                "s3:ListBucketVersions",
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::amzn-s3-demo-bucket",
                "arn:aws:s3:::amzn-s3-demo-bucket/*"
            ]
        }
    ]
}
```

## Políticas para conceder los permisos mínimos

La siguiente política otorga los permisos mínimos necesarios para utilizar Redshift Spectrum con Amazon S3, AWS Glue, y Athena.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "s3:ListBucketVersions",
                "s3:ListBucket",
                "s3:GetObject"
            ],
            "Resource": [
```

```
                "arn:aws:s3:::bucketname",
                "arn:aws:s3:::bucketname/folder1/folder2/*"
             ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "glue:CreateDatabase",
                "glue:DeleteDatabase",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:UpdateDatabase",
                "glue:CreateTable",
                "glue:DeleteTable",
                "glue:BatchDeleteTable",
                "glue:UpdateTable",
                "glue:GetTable",
                "glue:GetTables",
                "glue:BatchCreatePartition",
                "glue:CreatePartition",
                "glue:DeletePartition",
                "glue:BatchDeletePartition",
                "glue:UpdatePartition",
                "glue:GetPartition",
                "glue:GetPartitions",
                "glue:BatchGetPartition"
            ],
            "Resource": "*"
        }
    ]
}
```

Si utiliza Athena con el catálogo de datos en lugar de AWS Glue, la política necesitará tener acceso pleno a Athena. La siguiente política concede acceso a los recursos de Athena. Si la base de datos externa está en un metastore de Hive, no es necesario que tenga acceso a Athena.

```
{
 "Version": "2012-10-17",
 "Statement": [{
  "Effect": "Allow",
  "Action": ["athena:*"],
  "Resource": ["*"]
 }]
```

```
}
```

## Encadenamiento de roles de IAM en Amazon Redshift Spectrum

Cuando adjunta un rol a su clúster, este puede asumir ese rol y, de esa manera, obtener acceso a Amazon S3, Athena y AWS Glue en su nombre. Si un rol que está asociado al clúster no tiene acceso a los recursos necesarios, puede encadenar otro rol, que posiblemente pertenezca a otra cuenta. El clúster asumirá de forma temporal el rol encadenado para obtener acceso a los datos. También puede conceder acceso entre cuentas mediante el encadenamiento de roles. Puede encadenar un máximo de 10 roles. Cada uno de los roles de la cadena asume el rol siguiente, hasta que el clúster asume el rol situado al final de la cadena.

Para encadenar roles, debe establecer una relación de confianza entre ellos. Un rol que asume otro rol debe tener una política de permisos que le permita asumir el rol especificado. A su vez, el rol que transfiere los permisos debe tener una política de confianza que le permita transferir sus permisos a otro rol. Para obtener más información, consulte Encadenamiento de roles de IAM en Amazon Redshift.

Ejecute el comando CREATE EXTERNAL SCHEMA para encadenar funciones incluyendo una lista de ARN de funciones separados por comas.

> **Note**
>
> La lista de roles encadenados no debe incluir espacios.

En el siguiente ejemplo, `MyRedshiftRole` está asociado al clúster. `MyRedshiftRole` adopta la función `AcmeData`, que pertenece a la cuenta `111122223333`.

```
create external schema acme from data catalog
database 'acmedb' region 'us-west-2'
iam_role 'arn:aws:iam::123456789012:role/MyRedshiftRole,arn:aws:iam::111122223333:role/
AcmeData';
```

## Control del acceso al catálogo de datos de AWS Glue

Su usa AWS Glue para el catálogo de datos, puede aplicar un control de acceso más preciso al catálogo de datos de AWS Glue con su política de IAM. Por ejemplo, es posible que desee exponer solo unas pocas bases de datos y tablas a un rol de IAM específico.

En las siguientes secciones se describen las políticas de IAM para diferentes niveles de acceso a los datos almacenados en el catálogo de datos de AWS Glue.

Temas


- Política para operaciones de bases de datos
- Política para operaciones de tablas
- Política para operaciones de partición


## Política para operaciones de bases de datos

Si desea otorgar a los usuarios permisos para ver y crear una base de datos, necesitarán derechos de acceso a la base de datos y el catálogo de datos de AWS Glue.

La siguiente consulta de ejemplo crea una base de datos.

```
CREATE EXTERNAL SCHEMA example_db
FROM DATA CATALOG DATABASE 'example_db' region 'us-west-2'
IAM_ROLE 'arn:aws:iam::redshift-account:role/spectrumrole'
CREATE EXTERNAL DATABASE IF NOT EXISTS
```

La siguiente política de IAM concede los permisos mínimos necesarios para la creación de una base de datos.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:GetDatabase",
                "glue:CreateDatabase"
            ],
            "Resource": [
                "arn:aws:glue:us-west-2:redshift-account:database/example_db",
                "arn:aws:glue:us-west-2:redshift-account:catalog"
            ]
        }
    ]
```

```
}
```

La siguiente consulta de ejemplo enumera las bases de datos actuales.

```
SELECT * FROM SVV_EXTERNAL_DATABASES WHERE
databasename = 'example_db1' or databasename = 'example_db2';
```

La siguiente política de IAM concede los permisos mínimos necesarios para enumerar las bases de datos actuales.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:GetDatabases"
            ],
            "Resource": [
                "arn:aws:glue:us-west-2:redshift-account:database/example_db1",
                "arn:aws:glue:us-west-2:redshift-account:database/example_db2",
                "arn:aws:glue:us-west-2:redshift-account:catalog"

            ]
        }
    ]
}
```

## Política para operaciones de tablas

Si desea otorgar a los usuarios permisos para ver, crear, eliminar, modificar, o realizar otras acciones en tablas, estos usuarios necesitarán varios tipos de acceso. Necesitan acceso a las tablas, las bases de datos a las que pertenecen y el catálogo.

La siguiente consulta de ejemplo crea una tabla externa.

```
CREATE EXTERNAL TABLE example_db.example_tbl0(
    col0 INT,
    col1 VARCHAR(255)
) PARTITIONED BY (part INT) STORED AS TEXTFILE
LOCATION 's3://test/s3/location/';
```

La siguiente política de IAM concede los permisos mínimos necesarios para crear una tabla externa.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:CreateTable"
            ],
            "Resource": [
                "arn:aws:glue:us-west-2:redshift-account:catalog",
                "arn:aws:glue:us-west-2:redshift-account:database/example_db",
                "arn:aws:glue:us-west-2:redshift-account:table/example_db/example_tbl0"
            ]
        }
    ]
}
```

En el siguiente ejemplo, se consulta cada lista de las tablas externas actuales.

```
SELECT * FROM svv_external_tables
WHERE tablename = 'example_tbl0' OR
tablename = 'example_tbl1';
```

```
SELECT * FROM svv_external_columns
WHERE tablename = 'example_tbl0' OR
tablename = 'example_tbl1';
```

```
SELECT parameters FROM svv_external_tables
WHERE tablename = 'example_tbl0' OR
tablename = 'example_tbl1';
```

La siguiente política de IAM concede los permisos mínimos necesarios para enumerar las tablas externas actuales.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:GetTables"
            ],
            "Resource": [
                "arn:aws:glue:us-west-2:redshift-account:catalog",
                "arn:aws:glue:us-west-2:redshift-account:database/example_db",
                "arn:aws:glue:us-west-2:redshift-account:table/example_db/
example_tbl0",
                "arn:aws:glue:us-west-2:redshift-account:table/example_db/example_tbl1"

            ]
        }
    ]
}
```

La siguiente consulta de ejemplo modifica una tabla existente.

```
ALTER TABLE example_db.example_tbl0
SET TABLE PROPERTIES ('numRows' = '100');
```

La siguiente política de IAM concede los permisos mínimos necesarios para modificar una tabla existente.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:GetTable",
                "glue:UpdateTable"
            ],
            "Resource": [
                "arn:aws:glue:us-west-2:redshift-account:catalog",
                "arn:aws:glue:us-west-2:redshift-account:database/example_db",
                "arn:aws:glue:us-west-2:redshift-account:table/example_db/example_tbl0"

            ]
        }
    ]
}
```

La siguiente consulta de ejemplo elimina una tabla existente.

```
DROP TABLE example_db.example_tbl0;
```

La siguiente política de IAM concede los permisos mínimos necesarios para eliminar una tabla existente.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:DeleteTable"
            ],
            "Resource": [
                "arn:aws:glue:us-west-2:redshift-account:catalog",
                "arn:aws:glue:us-west-2:redshift-account:database/example_db",
                "arn:aws:glue:us-west-2:redshift-account:table/example_db/example_tbl0"
```

```
            ]
        }
    ]
}
```

## Política para operaciones de partición

Si desea otorgar a los usuarios permisos para realizar operaciones en el nivel de partición (ver, crear, eliminar, modificar, etc.), estos usuarios necesitarán permisos para las tablas a las que pertenecen las particiones. También necesitarán permisos para las bases de datos relacionadas y el catálogo de datos de AWS Glue.

La siguiente consulta de ejemplo crea una partición.

```
ALTER TABLE example_db.example_tbl0
ADD PARTITION (part=0) LOCATION 's3://test/s3/location/part=0/';
ALTER TABLE example_db.example_t
ADD PARTITION (part=1) LOCATION 's3://test/s3/location/part=1/';
```

La siguiente política de IAM concede los permisos mínimos necesarios para crear una partición.

```
 {
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:GetTable",
                "glue:BatchCreatePartition"
            ],
            "Resource": [
                "arn:aws:glue:us-west-2:redshift-account:catalog",
                "arn:aws:glue:us-west-2:redshift-account:database/example_db",
                "arn:aws:glue:us-west-2:redshift-account:table/example_db/example_tbl0"
            ]
        }
    ]
}
```

La siguiente consulta de ejemplo enumera las particiones actuales.

```
SELECT * FROM svv_external_partitions
WHERE schemname = 'example_db' AND
tablename = 'example_tbl0'
```

La siguiente política de IAM concede los permisos mínimos necesarios para enumerar las particiones actuales.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:GetPartitions",
                "glue:GetTables",
                "glue:GetTable"
            ],
            "Resource": [
                "arn:aws:glue:us-west-2:redshift-account:catalog",
                "arn:aws:glue:us-west-2:redshift-account:database/example_db",
                "arn:aws:glue:us-west-2:redshift-account:table/example_db/example_tbl0"
            ]
        }
    ]
}
```

La siguiente consulta de ejemplo modifica una partición existente.

```
ALTER TABLE example_db.example_tbl0 PARTITION(part='0')
SET LOCATION 's3://test/s3/new/location/part=0/';
```

La siguiente política de IAM concede los permisos mínimos necesarios para modificar una partición existente.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:GetPartition",
                "glue:UpdatePartition"
            ],
            "Resource": [
                "arn:aws:glue:us-west-2:redshift-account:catalog",
                "arn:aws:glue:us-west-2:redshift-account:database/example_db",
                "arn:aws:glue:us-west-2:redshift-account:table/example_db/example_tbl0"
            ]
        }
    ]
}
```

La siguiente consulta de ejemplo elimina una partición existente.

```
ALTER TABLE example_db.example_tbl0 DROP PARTITION(part='0');
```

La siguiente política de IAM concede los permisos mínimos necesarios para eliminar una partición existente.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
```

```
            "Action": [
                "glue:DeletePartition"
            ],
            "Resource": [
                "arn:aws:glue:us-west-2:redshift-account:catalog",
                "arn:aws:glue:us-west-2:redshift-account:database/example_db",
                "arn:aws:glue:us-west-2:redshift-account:table/example_db/example_tbl0"
            ]
        }
    ]
}
```

# Redshift Spectrum y AWS Lake Formation

En este tema se describe cómo utilizar Redshift Spectrum con Lake Formation. Lake Formation es un servicio para compartir datos de análisis.

Puede usar AWS Lake Formation para definir y aplicar de forma centralizada políticas de acceso en el nivel de columnas, bases de datos y tablas en los datos almacenados en Amazon S3. Después de que sus datos se registren en un AWS Glue Data Catalog habilitado para Lake Formation, puede consultarlos por medio de diversos servicios, incluido Redshift Spectrum.

Lake Formation proporciona la seguridad y la gobernanza del catálogo de datos. En Lake Formation, puede otorgar y revocar permisos a los objetos del catálogo de datos, como las bases de datos, las tablas, las columnas y el almacenamiento subyacente de Amazon S3.

> **Important**
>
> Solo se puede utilizar Redshift Spectrum con un catálogo de datos habilitado para Lake Formation en aquellas regiones de AWS en las que esté disponible Lake Formation. Para obtener una lista de las regiones disponibles, consulte Puntos de conexión y cuotas de AWS Lake Formation en la Referencia general de AWS.

Mediante el uso de Redshift Spectrum con Lake Formation, puede hacer lo siguiente:

- Utilice Lake Formation como lugar centralizado en donde conceda y revoque permisos y políticas de control del acceso sobre todos sus datos en el lago de datos. Lake Formation proporciona una jerarquía de permisos que controlan el acceso a las bases de datos y a las tablas de un catálogo

de datos. Para obtener más información, consulte Descripción general de permisos de Lake Formation en la Guía para desarrolladores de AWS Lake Formation.

- Cree tablas externas y realice búsquedas en los datos del lago de datos. Antes de que los usuarios en su cuenta puedan ejecutar consultas, un administrador de cuenta del lago de datos registra sus rutas existentes de Amazon S3 que contienen datos de la fuente con Lake Formation. El administrador también crea tablas y otorga permisos a sus usuarios. Se puede otorgar el acceso en bases de datos, tablas o columnas. El administrador puede usar filtros de datos en Lake Formation para garantizar un control de acceso granular sobre los datos confidenciales almacenados en Amazon S3. Para obtener más información, consulte Uso de filtros de datos para seguridad en el nivel de fila y de celda.

  Una vez registrados los datos en el catálogo de datos, cada vez que los usuarios intentan ejecutar consultas, Lake Formation verifica el acceso a la tabla para esa entidad principal específica. Lake Formation provee credenciales temporales a Redshift Spectrum, y se ejecuta la consulta.
- Ejecute consultas de Redshift Spectrum en función de un AWS Glue Data Catalog montado automáticamente mediante las credenciales de IAM obtenidas con `GetCredentials` o `GetClusterCredentials` y gestione los permisos de Lake Formation por usuario de la base de datos (IAMR:Username o IAM:Username).

Cuando utilice Redshift Spectrum con un catálogo de datos habilitado para Lake Formation, deberá cumplir uno de los siguientes requisitos:

- Un rol de IAM asociado al clúster que tiene permiso en el catálogo de datos.
- Una identidad de IAM federada configurada para administrar el acceso a recursos externos. Para obtener más información, consulte Uso de una identidad federada para administrar el acceso de Amazon Redshift a los recursos locales y a las tablas externas de Amazon Redshift.

> **Important**
>
> No es posible encadenar los roles de IAM mientras se utiliza Redshift Spectrum con un catálogo de datos habilitado para Lake Formation.

Para obtener más información acerca de los pasos necesarios para configurar AWS Lake Formation y así poder utilizarlo con Redshift Spectrum, consulte Tutorial: Creación de un lago de datos a partir de un origen JDBC en Lake Formation en la Guía para desarrolladores de AWS Lake Formation.

Para obtener más información sobre la integración con Redshift Spectrum, consulte Consulte los datos del lago de datos mediante Amazon Redshift Spectrum. Los datos y los recursos de AWS utilizados en este tema dependen de los pasos realizados con anterioridad en el tutorial.

## Uso de filtros de datos para seguridad en el nivel de fila y de celda

Puede definir filtros de datos en AWS Lake Formation para controlar el acceso a nivel de filas y celdas de sus consultas de Redshift Spectrum a los datos definidos en su catálogo de datos. Para configurarlo, realice las tareas siguientes:

- Cree un filtro de datos en Lake Formation con la siguiente información:
  - Especificación de columna con una lista de columnas para incluir o excluir de los resultados de la consulta.
  - Expresión de filtro de filas que especifica las filas que se van a incluir en los resultados de la consulta.

  Para obtener más información acerca de cómo crear un filtro de datos, consulte Filtros de datos en Lake Formation en la Guía para desarrolladores de AWS Lake Formation.
- Cree una tabla externa en Amazon Redshift que haga referencia a una tabla del catálogo de datos habilitado para Lake Formation. Para obtener más información sobre cómo consultar una tabla de Lake Formation con Redshift Spectrum, consulte Consulte los datos del lago de datos mediante Amazon Redshift Spectrum en la Guía para desarrolladores de AWS Lake Formation.

Una vez definida la tabla en Amazon Redshift, puede consultar la tabla Lake Formation y acceder únicamente a las filas y columnas permitidas por el filtro de datos.

Para obtener una guía detallada sobre cómo configurar la seguridad a nivel de fila y de celda en Lake Formation y, a continuación, consultar con Redshift Spectrum, consulte Use Amazon Redshift Spectrum with row-level and cell-level security policies defined in AWS Lake Formation.

# Archivos de datos para consultas en Amazon Redshift Spectrum

En esta sección se describe cómo crear archivos de datos en Amazon S3 en un formato compatible con Redshift Spectrum.

Los archivos de datos que utiliza para las consultas en Amazon Redshift Spectrum son, por lo general, los mismos tipos de archivos que usa para otras aplicaciones. Por ejemplo, se utilizan los mismos tipos de archivos con Amazon Athena, Amazon EMR y QuickSight. Puede consultar los

datos en su formato original desde Amazon S3 directamente. Para hacer esto, los archivos de datos deben estar en un formato que Redshift Spectrum admita y deben estar ubicados en un bucket de Amazon S3 al cual su clúster tenga acceso.

El bucket de Amazon S3 con los archivos de datos y el clúster de Amazon Redshift deben estar en la misma región de AWS. Para obtener información acerca de las regiones de AWS admitidas, consulte Regiones de Amazon Redshift Spectrum.

## Formatos de datos para Redshift Spectrum

Redshift Spectrum admite los siguientes formatos de datos estructurados y semiestructurados.

| Formato de archivo | En columnas | Admite lecturas paralelas | Unidad dividida |
|---|---|---|---|
| Parquet | Sí | Sí | Grupo de filas |
| ORC | Sí | Sí | Stripe |
| RCFile | Sí | Sí | Grupo de filas |
| TextFile | No | Sí | Fila |
| SequenceFile | No | Sí | Fila o bloque |
| RegexSerde | No | Sí | Fila |
| OpenCSV | No | Sí | Fila |
| AVRO | No | Sí | Bloque |
| Ion | No | No | N/A |
| JSON | No | No | N/A |

En la tabla anterior, los encabezados indican lo siguiente:

- En columnas: si el formato de archivo almacena físicamente los datos en una estructura orientada a columnas en lugar de hacerlo en una orientada a filas.

- Admite lecturas paralelas: si el formato de archivo admite la lectura de bloques individuales dentro del archivo. La lectura de bloques individuales permite el procesamiento distribuido de un archivo entre varias solicitudes de Redshift Spectrum independientes en lugar de tener que leer el archivo completo en una sola solicitud.
- Unidad dividida: para formatos de archivo que se pueden leer en paralelo, la unidad dividida es el fragmento más pequeño de datos que puede procesar una sola solicitud de Redshift Spectrum.

> **Note**
>
> Los valores de marca temporal de los archivos de texto deben tener el formato `yyyy-MM-dd HH:mm:ss.SSSSSS`, tal y como se muestra en el siguiente valor de marca temporal: `2017-05-01 11:30:59.000000`.

Recomendamos utilizar un formato de archivo de almacenamiento en columnas, como Apache Parquet. Amazon Redshift admite el formato de archivo de datos Apache Parquet v1. Con un formato de archivo de almacenamiento en columnas, se puede reducir la transferencia de datos desde Amazon S3 al seleccionar únicamente las columnas que necesite.

## Tipos de compresión para Redshift Spectrum

Para reducir el espacio de almacenamiento, mejorar el rendimiento y reducir los costos, se recomienda encarecidamente comprimir los archivos de datos. Redshift Spectrum reconoce los tipos de compresión de archivos basados en la extensión del archivo.

Redshift Spectrum admite los siguientes tipos de compresión y extensiones.

| Algoritmo de compresión | Extensión de archivo | Soporta lecturas paralelas |
|---|---|---|
| Gzip | .gz | No |
| Bzip2 | .bz2 | Sí |
| Snappy | .snappy | No |

Puede aplicar compresión a distintos niveles. En la mayoría de los casos, se comprime un archivo completo o se comprimen bloques individuales dentro de un archivo. La compresión de los formatos en columna en el nivel de archivo no ofrece beneficios en el rendimiento.

Para que Redshift Spectrum pueda leer un archivo en paralelo, debe cumplirse lo siguiente:

- El formato de archivo admite lecturas paralelas.
- La compresión de nivel de archivo, si la hay, admite lecturas paralelas.

No importa si las unidades divididas individuales dentro de un archivo se comprimen con un algoritmo de compresión que se puede leer en paralelo, ya que cada unidad dividida se procesa con una sola solicitud de Redshift Spectrum. Un ejemplo de esto son los archivos Parquet comprimidos con Snappy. Los grupos de filas individuales del archivo Parquet se comprimen con Snappy, pero la estructura de nivel superior del archivo permanece sin comprimir. En este caso, el archivo se puede leer en paralelo porque cada solicitud de Redshift Spectrum puede leer y procesar grupos de filas individuales de Amazon S3.

## Cifrado para Redshift Spectrum

Redshift Spectrum descifra de manera transparente archivos de datos cifrados que hayan utilizando las siguientes opciones de cifrado:

- cifrado del lado del servidor (SSE-S3) a través de una clave de cifrado AES-256 administrada por Amazon S3
- cifrado del lado del servidor con claves administradas por AWS Key Management Service (SSE-KMS)

Redshift Spectrum no admite el cifrado del lado del cliente de Amazon S3. Para obtener más información sobre el cifrado del lado del servidor, consulte Protección de datos con el cifrado del lado del servidor en la Guía del usuario de Amazon Simple Storage Service.

Amazon Redshift utiliza un procesamiento masivo en paralelo (MPP) para lograr una ejecución rápida de las consultas complejas que operan sobre grandes cantidades de datos. Redshift Spectrum extiende el mismo principio para consultar datos externos, utilizando distintas instancias de Redshift Spectrum según se necesite para examinar archivos. Coloque los archivos en una carpeta separada para cada tabla.

Puede optimizar los datos para el procesamiento en paralelo haciendo lo siguiente:

- Si el formato de archivo o la compresión no admiten la lectura en paralelo, divida los archivos grandes en muchos archivos más pequeños. Recomendamos utilizar tamaños de archivo de entre 64 MB y 1 GB.
- Mantenga todos los archivos con, aproximadamente, el mismo tamaño. Si algunos archivos son mucho más grandes que otros, Redshift Spectrum no podrá distribuir uniformemente la carga de trabajo.

# Esquemas externos en Amazon Redshift Spectrum

En este tema se describe cómo crear y usar esquemas externos con Redshift Spectrum. Los esquemas externos son conjuntos de tablas que se utilizan como referencias para acceder a datos fuera del clúster de Amazon Redshift. Estas tablas contienen metadatos sobre los datos externos que lee Redshift Spectrum.

Todas las tablas externas deben crearse en un esquema externo, que puede crear utilizando una instrucción CREATE EXTERNAL SCHEMA.

> **Note**
>
> Algunas aplicaciones utilizan indistintamente el término base de datos y esquema. En Amazon Redshift, se utiliza el término esquema.

Un esquema externo de Amazon Redshift referencia una base de datos externa en un catálogo de datos externo. Puede crear la base de datos externa en Amazon Redshift, Amazon Athena, AWS Glue Data Catalog o un metastore de Apache Hive, como Amazon EMR. Si crea una base de datos externa en Amazon Redshift, la base de datos se encuentra en el catálogo de datos de Athena. Para crear una base de datos en un metaalmacén Hive, necesita crear la base de datos en su aplicación Hive.

Amazon Redshift necesita autorización para obtener acceso al catálogo de datos de Athena y a los archivos de datos de Amazon S3 en su nombre. Para proporcionar esa autorización, primero cree un rol de AWS Identity and Access Management (IAM). Después tiene que adjuntar el rol al clúster y proporcionar el nombre de recurso de Amazon (ARN) para ese rol en la instrucción `CREATE EXTERNAL SCHEMA` de Amazon Redshift. Para obtener más información acerca de la autorización, consulte Políticas de IAM para Amazon Redshift Spectrum.

Para crear una base de datos externa en el mismo momento en que crea un esquema externo, especifique el `FROM DATA CATALOG` e incluya la cláusula `CREATE EXTERNAL DATABASE` en su instrucción `CREATE EXTERNAL SCHEMA`.

En el siguiente ejemplo, se crea un esquema externo denominado `spectrum_schema` utilizando la base de datos externa `spectrum_db`.

```
create external schema spectrum_schema from data catalog
database 'spectrum_db'
iam_role 'arn:aws:iam::123456789012:role/MySpectrumRole'
create external database if not exists;
```

Si administra un catálogo de datos con Athena, especifique el nombre de la base de datos de Athena y la región de AWS en la que se encuentra el catálogo de datos de Athena.

En el siguiente ejemplo, se crea un esquema externo mediante la base de datos `sampledb` predeterminada en el catálogo de datos de Athena.

```
create external schema athena_schema from data catalog
database 'sampledb'
iam_role 'arn:aws:iam::123456789012:role/MySpectrumRole'
region 'us-east-2';
```

> **Note**
>
> El parámetro `region` referencia la región de AWS en la que se encuentra el catálogo de datos de Athena y no a la ubicación de los archivos de datos de Amazon S3.

Si administra el catálogo de datos mediante un metastore de Hive, como Amazon EMR, sus grupos de seguridad se deben configurar de manera tal que permitan el tráfico entre los clústeres.

En la instrucción CREATE EXTERNAL SCHEMA, especifique la opción `FROM HIVE METASTORE` e incluya el Uniform Resource Identifier (URI, Identificador uniforme de recursos) y el número de puerto del metaalmacén. En el siguiente ejemplo, se crea un esquema externo a través de una base de datos de metaalmacén Hive denominada `hive_db`.

```
create external schema hive_schema
from hive metastore
```

```
database 'hive_db'
uri '172.10.10.10' port 99
iam_role 'arn:aws:iam::123456789012:role/MySpectrumRole'
```

Para ver los esquemas externos para su clúster, consulte la tabla de catálogo PG_EXTERNAL_SCHEMA o la vista SVV_EXTERNAL_SCHEMAS. En el siguiente ejemplo, se consulta SVV_EXTERNAL_SCHEMAS, que combina PG_EXTERNAL_SCHEMA y PG_NAMESPACE.

```
select * from svv_external_schemas
```

Para ver la sintaxis completa del comando y ejemplos, consulte CREATE EXTERNAL SCHEMA.

## Uso de catálogos externos en Amazon Redshift Spectrum

Los metadatos para las bases de datos y las tablas externas de Amazon Redshift Spectrum se almacenan en un catálogo de datos externo. De manera predeterminada, los metadatos de Redshift Spectrum se almacenan en un catálogo de datos de Athena. Puede ver y administrar bases de datos y tablas de Redshift Spectrum en su consola de Athena.

También puede crear y administrar bases de datos y tablas externas mediante el lenguaje de definición de datos (DDL) de Hive a través de Athena o un metastore de Hive, como Amazon EMR.

> **Note**
>
> Se recomienda que utilice Amazon Redshift para crear y administrar las bases de datos y las tablas externas de Redshift Spectrum.

### Visualización de las bases de datos de Redshift Spectrum en Athena y AWS Glue

Puede crear una base de datos externa que incluye la cláusula CREATE EXTERNAL DATABASE IF NOT EXISTS como parte de su instrucción CREATE EXTERNAL SCHEMA. En estos casos, los metadatos de la base de datos externa se almacenan en su catálogo de datos. Los metadatos para las tablas externas que cree calificadas por el esquema externo también se almacenan en su catálogo de datos de .

Athena y AWS Glue mantienen un catálogo de datos por cada Región de AWS admitida. Para ver los metadatos de la tabla, inicie sesión en Athena o la consola de AWS Glue. En Athena, elija Data

sources (Orígenes de datos), su AWS Glue y, a continuación, vea los detalles de su base de datos. En AWS Glue, elija Databases (Bases de datos), su base de datos externa y luego vea los detalles de su base de datos.

Si crea y administra las tablas externas con Athena, registre la base de datos con CREATE EXTERNAL SCHEMA. Por ejemplo, el siguiente comando registra la base de datos de Athena denominada `sampledb`.

```
create external schema athena_sample
from data catalog
database 'sampledb'
iam_role 'arn:aws:iam::123456789012:role/mySpectrumRole'
region 'us-east-1';
```

Cuando consulta la vista de sistema SVV_EXTERNAL_TABLES, ve las tablas de la base de datos `sampledb` de Athena y también las tablas que creó en Amazon Redshift.

```
select * from svv_external_tables;
```

```
schemaname    | tablename        | location

--------------+------------------
+-----------------------------------------------------
athena_sample | elb_logs         | s3://athena-examples/elb/plaintext
athena_sample | lineitem_1t_csv  | s3://myspectrum/tpch/1000/lineitem_csv

athena_sample | lineitem_1t_part | s3://myspectrum/tpch/1000/lineitem_partition

spectrum      | sales            | s3://redshift-downloads/tickit/spectrum/sales

spectrum      | sales_part       | s3://redshift-downloads/tickit/spectrum/sales_part
```

## Registro de una base de datos de metaalmacén Apache Hive

Si crea tablas externas en un metaalmacén Apache Hive, puede usar CREATE EXTERNAL SCHEMA para registrar esas tablas en Redshift Spectrum.

En la instrucción CREATE EXTERNAL SCHEMA, especifique la cláusula FROM HIVE METASTORE y brinde el Uniform Resource Identifier (URI, Identificador uniforme de recursos) y el número de puerto del metaalmacén Hive. El rol de IAM debe incluir un permiso de acceso para Amazon S3,

aunque no necesita permisos para Athena. En el siguiente ejemplo, se registra un metaalmacén Hive.

```
create external schema if not exists hive_schema
from hive metastore
database 'hive_database'
uri 'ip-10-0-111-111.us-west-2.compute.internal' port 9083
iam_role 'arn:aws:iam::123456789012:role/mySpectrumRole';
```

## Habilitación del acceso del clúster de Amazon Redshift al clúster de Amazon EMR

Si el metastore de Hive está en Amazon EMR, debe conceder al clúster de Amazon Redshift acceso a su clúster de Amazon EMR. Para ello, crea un grupo de seguridad de Amazon EC2. A continuación, permite todo el tráfico entrante al grupo de seguridad de EC2 procedente del grupo de seguridad del clúster de Amazon Redshift y del grupo de seguridad del clúster de Amazon EMR. Luego, agrega la seguridad de EC2 tanto al clúster de Amazon Redshift como al de Amazon EMR.

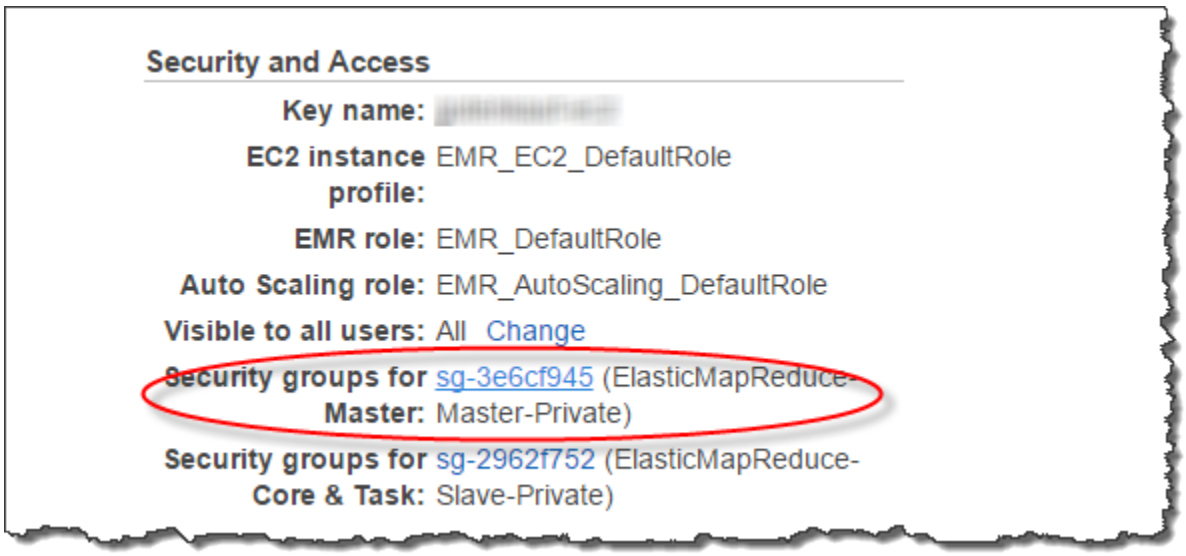
Visualización del nombre del grupo de seguridad de su clúster de Amazon Redshift

Para mostrar el grupo de seguridad, haga lo siguiente:

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. En el menú de navegación, elija Clusters (Clústeres) y, a continuación, elija el clúster de la lista para abrir sus detalles.
3. Elija Properties (Propiedades) y vea la sección Network and security (Configuración de redes y seguridad).
4. Busque su grupo de seguridad en VPC security group (Grupo de seguridad de la VPC) y anótelo.

Visualización del nombre del grupo de seguridad del nodo maestro de Amazon EMR

1. Abra su clúster de Amazon EMR. Para obtener más información, consulte Uso de configuraciones de seguridad para definir la seguridad del clúster en la Guía de administración de Amazon EMR.
2. En Security and access (Seguridad y acceso), anote el nombre del grupo de seguridad del nodo maestro de Amazon EMR.




Para crear o modificar un grupo de seguridad de Amazon EC2 para permitir la conexión entre Amazon Redshift y Amazon EMR

1. En el panel de Amazon EC2, elija Security Groups (Grupos de seguridad). Para obtener más información, consulte Reglas del grupo de seguridad en la Guía del usuario de Amazon EC2.
2. Elija Create Security Group (Crear grupo de seguridad).
3. Si utiliza una VPC, elija la VPC en la que estén el clúster de Amazon Redshift y el de Amazon EMR.
4. Agregue una regla de entrada.
    1. En Type (Tipo), elija Custom TCP (TCP personalizada).
    2. En Source (Origen), seleccione Custom (Personalizado).
    3. Ingrese el nombre del grupo de seguridad de Amazon Redshift.
5. Agregue otra regla de entrada.
    1. En Type (Tipo), seleccione TCP.
    2. En Port Range (Rango de puertos), escriba 9083.

> **Note**
>
> El puerto predeterminado para EMR HMS es 9083. Si su HMS utiliza un puerto diferente, especifique ese puerto en la regla de entrada y en la definición del esquema externo.

3. En Source (Origen), seleccione Custom (Personalizado).

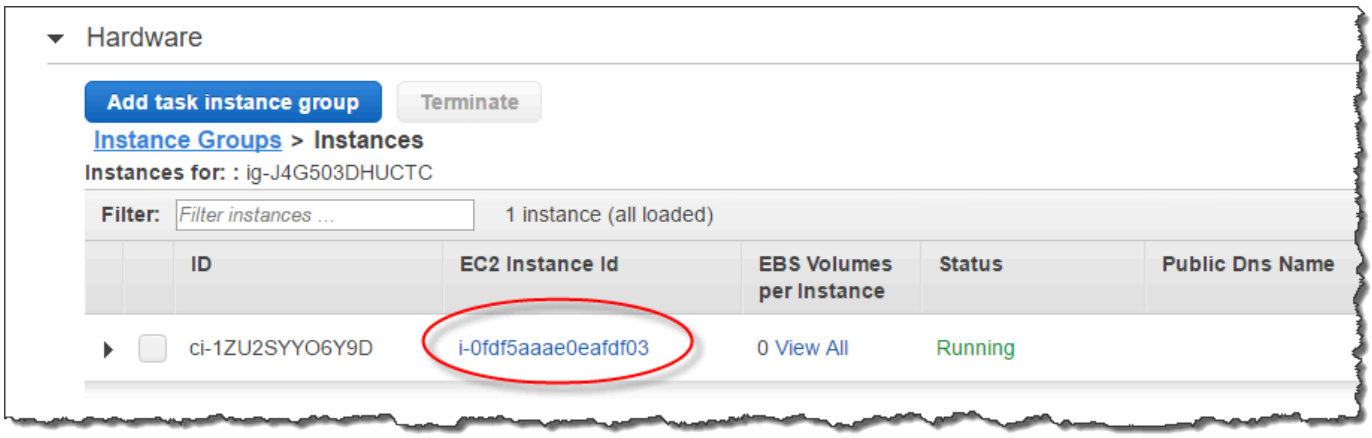
6. Ingrese un nombre y una descripción para el grupo de seguridad.
7. Elija Create Security Group (Crear grupo de seguridad).

Para agregar el grupo de seguridad de Amazon EC2 que creó en el procedimiento anterior al clúster de Amazon Redshift

1. En Amazon Redshift, elija el clúster.
2. Seleccione Properties (Propiedades).
3. Vea la sección Network and security settings (Configuración de redes y seguridad) y, a continuación, elija Edit (Editar).
4. En la opción VPC security group (Grupo de seguridad de la VPC), elija el nuevo nombre del grupo de seguridad.
5. Elija Save changes (Guardar cambios).

Para agregar el grupo de seguridad de Amazon EC2 a su clúster de Amazon EMR

1. En Amazon EMR, elija el clúster. Para obtener más información, consulte Uso de configuraciones de seguridad para definir la seguridad del clúster en la Guía de administración de Amazon EMR.
2. En Hardware, seleccione el enlace del nodo principal.
3. Elija el enlace en la columna EC2 Instance ID (ID de instancia EC2).




4. En Actions (Acciones), elija Security (Seguridad) y, luego, Change security groups (Cambiar grupos de seguridad).
5. En Associated sercurity groups (Grupos de seguridad asociados), elija el nuevo grupo de seguridad y, luego, Add security group (Agregar grupo de seguridad).
6. Seleccione Save.

# Tablas externas para Redshift Spectrum

En este tema se describe cómo crear y usar tablas externas con Redshift Spectrum. Las tablas externas son tablas que se utilizan como referencias para acceder a datos fuera del clúster de Amazon Redshift. Estas tablas contienen metadatos sobre los datos externos que lee Redshift Spectrum.

La tabla externa se crea en un esquema externo. Para crear tablas externas, debe ser el propietario del esquema externo o un superusuario. Para transferir la propiedad de un esquema externo, use ALTER SCHEMA para cambiar el propietario. En el siguiente ejemplo, se modifica el propietario del esquema `spectrum_schema` a `newowner`.

```
alter schema spectrum_schema owner to newowner;
```

Para ejecutar una consulta de Redshift Spectrum, necesita los siguientes permisos:

- Permiso de uso para el esquema
- Permiso para crear tablas temporales en la base de datos actual

En el siguiente ejemplo, se concede permiso de uso para el esquema `spectrum_schema` al grupo de usuarios `spectrumusers`.

```
grant usage on schema spectrum_schema to group spectrumusers;
```

En el siguiente ejemplo, se concede un permiso temporal para la base de datos `spectrumdb` al grupo de usuarios `spectrumusers`.

```
grant temp on database spectrumdb to group spectrumusers;
```

Puede crear una tabla externa en Amazon Redshift, AWS Glue, Amazon Athena o un metastore de Apache Hive. Para obtener más información, consulte Introducción al uso de AWS Glue en la Guía para desarrolladores de AWS Glue, Introducción en la Guía del usuario de Amazon Athena, o bien, Apache Hive en la Guía para desarrolladores de Amazon EMR.

Si la tabla externa se define en AWS Glue, Athena o un metastore de Hive, primero debe crear un esquema externo que referencie la base de datos externa. A continuación, puede referenciar la tabla externa en la instrucción SELECT al colocar como prefijo el nombre del esquema en el nombre de la tabla, sin que sea necesario crear la tabla en Amazon Redshift. Para obtener más información, consulte Esquemas externos en Amazon Redshift Spectrum.

Para permitir que Amazon Redshift visualice las tablas en el AWS Glue Data Catalog, agregue `glue:GetTable` al rol de IAM de Amazon Redshift. De lo contrario, es posible que obtenga un error similar al siguiente.

```
RedshiftIamRoleSession is not authorized to perform: glue:GetTable on resource: *;
```

Por ejemplo, supongamos que tiene una tabla externa denominada `lineitem_athena` que está definida en un catálogo externo de Athena. En este caso, puede definir un esquema externo denominado `athena_schema` y, luego, consultar la tabla utilizando la siguiente instrucción SELECT.

```
select count(*) from athena_schema.lineitem_athena;
```

Para definir una tabla externa en Amazon Redshift, utilice el comando CREATE EXTERNAL TABLE. La instrucción de tabla externa define las columnas de la tabla, el formato de los archivos de datos y la ubicación de los datos en Amazon S3. Redshift Spectrum examina los archivos en la carpeta especificada y en cualquier subcarpeta. Redshift Spectrum ignora los archivos ocultos y los archivos

que comienzan con un punto, un guion bajo o una marca hash ( . , _, o #) o que finalizan con una tilde (~).

En el siguiente ejemplo, se crea una tabla denominada SALES en el esquema externo `spectrum` de Amazon Redshift. Los datos están en archivos de texto delimitados por tabulaciones.

```
create external table spectrum.sales(
salesid integer,
listid integer,
sellerid integer,
buyerid integer,
eventid integer,
dateid smallint,
qtysold smallint,
pricepaid decimal(8,2),
commission decimal(8,2),
saletime timestamp)
row format delimited
fields terminated by '\t'
stored as textfile
location 's3://redshift-downloads/tickit/spectrum/sales/'
table properties ('numRows'='172000');
```

Para ver las tablas externas, consulte la vista del sistema SVV_EXTERNAL_TABLES.

## Pseudocolumnas

De forma predeterminada, Amazon Redshift crea tablas externas con las pseudocolumnas `$path`, `$size` y `$spectrum_oid`. Seleccione la columna `$path` para ver la ruta a los archivos de datos en Simple Storage Service (Amazon S3) y seleccione la columna `$size` para ver el tamaño de los archivos de datos de cada fila devuelta por una consulta. La columna `$spectrum_oid` proporciona la capacidad de realizar consultas correlacionadas con Redshift Spectrum. Para ver un ejemplo, consulta Ejemplo: Realización de subconsultas correlacionadas en Redshift Spectrum. Debe delimitar los nombres de las columnas `$path`, `$size` y `$spectrum_oid` con comillas dobles. Una cláusula SELECT * no devuelve las pseudocolumnas. Debe incluir explícitamente los nombres de columna `$path`, `$size` y `$spectrum_oid` en la consulta, tal y como se muestra en el siguiente ejemplo.

```
select "$path", "$size", "$spectrum_oid"
from spectrum.sales_part where saledate = '2008-12-01';
```

Puede desactivar la creación de pseudocolumnas para una sesión estableciendo el parámetro de configuración `spectrum_enable_pseudo_columns` en `false`. Para obtener más información, consulte spectrum_enable_pseudo_columns. También puede desactivar solo la pseudocolumna `$spectrum_oid` configurando `enable_spectrum_oid` en `false`. Para obtener más información, consulte enable_spectrum_oid. Sin embargo, desactivar la pseudocolumna `$spectrum_oid` también desactiva la compatibilidad con consultas correlacionadas con Redshift Spectrum.

> **Important**
>
> Si se selecciona `$size`, `$path` o `$spectrum_oid`, se aplicarán cargos, ya que Redshift Spectrum analiza los archivos de datos de Simple Storage Service (Amazon S3) para determinar el tamaño del conjunto de resultados. Para obtener más información, consulte Precios de Amazon Redshift.

### Ejemplo de pseudocolumnas

El siguiente ejemplo devuelve el tamaño total de los archivos de datos relacionados de una tabla externa.

```
select distinct "$path", "$size"
from spectrum.sales_part;

 $path                                                                      | $size
----------------------------------------------------------------------------+-------
s3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-01/ |  1616
s3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-02/ |  1444
s3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-03/ |  1644
```

## Partición de tablas externas de Redshift Spectrum

Cuando particiona datos, puede restringir la cantidad de datos que Redshift Spectrum examina al filtrar mediante una clave de partición. Puede particionar datos por cualquier clave.

Una práctica habitual es particionar datos según el tiempo. Por ejemplo, puede seleccionar particionar por año, mes, fecha y hora. Si tiene datos que vienen de distintos orígenes, puede particionar por identificador de origen de datos y fecha.

En el siguiente procedimiento, se describe cómo particionar datos.

Pasos para particionar datos

1. Almacene los datos en carpetas en Amazon S3 según la clave de partición.

   Cree una carpeta para cada valor de partición y nombre esa carpeta con la clave de partición y el valor. Por ejemplo, si particiona por fecha, puede tener carpetas que se llamen `saledate=2017-04-01`, `saledate=2017-04-02` y, así, sucesivamente. Redshift Spectrum examina los archivos en la carpeta de la partición y en todas las subcarpetas. Redshift Spectrum ignora los archivos ocultos y los archivos que comienzan con un punto, un guion bajo o una marca hash ( . , _, o #) o que finalizan con una tilde (~).

2. Cree una tabla externa y especifique la clave de partición en la cláusula PARTITIONED BY.

   La clave de partición no puede ser el nombre de una columna de tabla. Los tipos de datos pueden ser SMALLINT, INTEGER, BIGINT, DECIMAL, REAL, DOUBLE PRECISION, BOOLEAN, CHAR, VARCHAR, DATE o TIMESTAMP.

3. Agregue las particiones.

   Agregue cada una de las particiones con ALTER TABLE … ADD PARTITION y especifique el valor de clave y la columna de partición, así como la ubicación de la carpeta de partición en Amazon S3. Puede agregar múltiples particiones en una única instrucción ALTER TABLE … ADD. En el siguiente ejemplo, se agregan particiones para `'2008-01'` y `'2008-03'`.

```
alter table spectrum.sales_part add
partition(saledate='2008-01-01')
location 's3://redshift-downloads/tickit/spectrum/sales_partition/
saledate=2008-01/'
partition(saledate='2008-03-01')
location 's3://redshift-downloads/tickit/spectrum/sales_partition/
saledate=2008-03/';
```

> **Note**
>
> Si usa el catálogo de AWS Glue, puede agregar hasta 100 particiones mediante una única instrucción ALTER TABLE.

## Ejemplos de partición de datos

En este ejemplo, se crea una tabla externa que está particionada por una sola clave de partición y una tabla externa que está particionada por dos claves de partición.

Los datos de muestra para este ejemplo se encuentran en un bucket de Amazon S3 que brinda acceso de lectura a todos los usuarios autenticados de AWS. El clúster y los archivos de datos externos deben estar en la misma Región de AWS. El bucket de datos de muestra está en la región Este de EE. UU. (Norte de Virginia) (us-east-1). Para obtener acceso a los datos utilizando Redshift Spectrum, el clúster también debe estar en us-east-2. Para que se muestren las carpetas en Amazon S3, ejecute los siguientes comandos.

```
aws s3 ls s3://redshift-downloads/tickit/spectrum/sales_partition/
```

```
PRE saledate=2008-01/
PRE saledate=2008-03/
PRE saledate=2008-04/
PRE saledate=2008-05/
PRE saledate=2008-06/
PRE saledate=2008-12/
```

Si aún no tiene un esquema externo, ejecute el siguiente comando. Sustituya el nombre de recurso de Amazon (ARN) para su rol de AWS Identity and Access Management (IAM).

```
create external schema spectrum
from data catalog
database 'spectrumdb'
iam_role 'arn:aws:iam::123456789012:role/myspectrumrole'
create external database if not exists;
```

### Ejemplo 1: Particionar con una sola clave de partición

En el siguiente ejemplo, se crea una tabla externa que se particiona por mes.

Para crear una tabla externa particionada por mes, ejecute el siguiente comando.

```
create external table spectrum.sales_part(
salesid integer,
listid integer,
sellerid integer,
```

```
buyerid integer,
eventid integer,
dateid smallint,
qtysold smallint,
pricepaid decimal(8,2),
commission decimal(8,2),
saletime timestamp)
partitioned by (saledate char(10))
row format delimited
fields terminated by '|'
stored as textfile
location 's3://redshift-downloads/tickit/spectrum/sales_partition/'
table properties ('numRows'='172000');
```

Para agregar las particiones, ejecute el siguiente comando ALTER TABLE.

```
alter table spectrum.sales_part add
partition(saledate='2008-01')
location 's3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-01/'

partition(saledate='2008-03')
location 's3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-03/'

partition(saledate='2008-04')
location 's3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-04/';
```

Para seleccionar los datos de la tabla particionada, ejecute la siguiente consulta.

```
select top 5 spectrum.sales_part.eventid, sum(spectrum.sales_part.pricepaid)
from spectrum.sales_part, event
where spectrum.sales_part.eventid = event.eventid
  and spectrum.sales_part.pricepaid > 30
  and saledate = '2008-01'
group by spectrum.sales_part.eventid
order by 2 desc;
```

```
eventid | sum
--------+---------
   4124 | 21179.00
   1924 | 20569.00
   2294 | 18830.00
   2260 | 17669.00
```

```
   6032 | 17265.00
```

Para ver particiones de la tabla externa, consulte la vista del sistema SVV_EXTERNAL_PARTITIONS.

```
select schemaname, tablename, values, location from svv_external_partitions
where tablename = 'sales_part';
```

```
schemaname | tablename  | values      | location

-----------+------------+-------------
+-------------------------------------------------------------------------
spectrum   | sales_part | ["2008-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-01
spectrum   | sales_part | ["2008-03"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-03
spectrum   | sales_part | ["2008-04"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-04
```

Ejemplo 2: Particionar con varias claves de partición

Para crear una tabla externa particionada por `date` y `eventid`, ejecute el siguiente comando.

```
create external table spectrum.sales_event(
salesid integer,
listid integer,
sellerid integer,
buyerid integer,
eventid integer,
dateid smallint,
qtysold smallint,
pricepaid decimal(8,2),
commission decimal(8,2),
saletime timestamp)
partitioned by (salesmonth char(10), event integer)
row format delimited
fields terminated by '|'
stored as textfile
location 's3://redshift-downloads/tickit/spectrum/salesevent/'
table properties ('numRows'='172000');
```

Para agregar las particiones, ejecute el siguiente comando ALTER TABLE.

```
alter table spectrum.sales_event add
partition(salesmonth='2008-01', event='101')
location 's3://redshift-downloads/tickit/spectrum/salesevent/salesmonth=2008-01/
event=101/'

partition(salesmonth='2008-01', event='102')
location 's3://redshift-downloads/tickit/spectrum/salesevent/salesmonth=2008-01/
event=102/'

partition(salesmonth='2008-01', event='103')
location 's3://redshift-downloads/tickit/spectrum/salesevent/salesmonth=2008-01/
event=103/'

partition(salesmonth='2008-02', event='101')
location 's3://redshift-downloads/tickit/spectrum/salesevent/salesmonth=2008-02/
event=101/'

partition(salesmonth='2008-02', event='102')
location 's3://redshift-downloads/tickit/spectrum/salesevent/salesmonth=2008-02/
event=102/'

partition(salesmonth='2008-02', event='103')
location 's3://redshift-downloads/tickit/spectrum/salesevent/salesmonth=2008-02/
event=103/'

partition(salesmonth='2008-03', event='101')
location 's3://redshift-downloads/tickit/spectrum/salesevent/salesmonth=2008-03/
event=101/'

partition(salesmonth='2008-03', event='102')
location 's3://redshift-downloads/tickit/spectrum/salesevent/salesmonth=2008-03/
event=102/'

partition(salesmonth='2008-03', event='103')
location 's3://redshift-downloads/tickit/spectrum/salesevent/salesmonth=2008-03/
event=103/';
```

Ejecute la siguiente consulta para seleccionar los datos de la tabla particionada.

```
select spectrum.sales_event.salesmonth, event.eventname,
 sum(spectrum.sales_event.pricepaid)
from spectrum.sales_event, event
where spectrum.sales_event.eventid = event.eventid
```

```
  and salesmonth = '2008-02'
 and (event = '101'
 or event = '102'
 or event = '103')
group by event.eventname, spectrum.sales_event.salesmonth
order by 3 desc;
```

```
salesmonth | eventname       | sum
-----------+-----------------+--------
2008-02    | The Magic Flute | 5062.00
2008-02    | La Sonnambula   | 3498.00
2008-02    | Die Walkure     |  534.00
```

## Asignación de columnas de tablas externas a columnas ORC

Las tablas externas de Amazon Redshift Spectrum se utilizan para consultar datos que están en archivos con formato ORC. El formato ORC (Optimized Row Columnar) es un archivo de almacenamiento organizado en columnas que admite estructuras de datos anidadas. Para obtener más información sobre la consulta de datos anidados, consulte Consultas de datos anidados con Amazon Redshift Spectrum.

Cuando crea una tabla externa que hace referencia a datos de un archivo ORC, debe asignar cada columna de la tabla externa a una columna de los datos de ORC. Para ello, puede utilizar uno de los siguientes métodos:

- Asignación por posición
- Asignación por nombre de columna

La asignación por nombre de columna es el valor predeterminado.

### Asignación por posición

En la asignación por posición, la primera columna definida en la tabla externa se asigna a la primera columna del archivo de datos de ORC y lo mismo ocurre con la segunda, la tercera, etc. En la asignación por posición, es necesario que el orden de las columnas de la tabla externa y del archivo ORC coincidan. Si el orden no coincide, puede asignar las columnas por nombre.

**Important**

En versiones anteriores, Redshift Spectrum utilizaba la asignación por posición de manera predeterminada. Si necesita seguir utilizando la asignación por posición en las tablas existentes, establezca la propiedad `orc.schema.resolution` de la tabla en `position`, tal y como se muestra en el siguiente ejemplo.

```
alter table spectrum.orc_example
set table properties('orc.schema.resolution'='position');
```

Por ejemplo, la tabla `SPECTRUM.ORC_EXAMPLE` se define del modo siguiente.

```
create external table spectrum.orc_example(
int_col int,
float_col float,
nested_col struct<
  "int_col" : int,
  "map_col" : map<int, array<float >>
   >
) stored as orc
location 's3://example/orc/files/';
```

La estructura de la tabla puede abstraerse tal y como se muestra a continuación.

```
• 'int_col' : int
• 'float_col' : float
• 'nested_col' : struct
   o 'int_col' : int
   o 'map_col' : map
      - key : int
      - value : array
         - value : float
```

El archivo ORC subyacente tiene la siguiente estructura de archivos.

```
• ORC file root(id = 0)
   o 'int_col' : int (id = 1)
   o 'float_col' : float (id = 2)
   o 'nested_col' : struct (id = 3)
```

```
      - 'int_col' : int (id = 4)
      - 'map_col' : map (id = 5)
         - key : int (id = 6)
         - value : array (id = 7)
            - value : float (id = 8)
```

En este ejemplo, puede asignar cada columna de la tabla externa a una columna del archivo ORC exclusivamente por posición. En el siguiente ejemplo, se muestra la asignación.

| Nombre de columna de la tabla externa | ID de columna ORC | Nombre de columna ORC |
|---|---|---|
| int_col | 1 | int_col |
| float_col | 2 | float_col |
| nested_col | 3 | nested_col |
| nested_col.int_col | 4 | int_col |
| nested_col.map_col | 5 | map_col |
| nested_col.map_col.key | 6 | N/D |
| nested_col.map_col.value | 7 | N/D |
| nested_col.map_col.value.item | 8 | N/D |

## Asignación por nombre de columna

En la asignación por nombre, las columnas de una tabla externa se asignan a columnas de los archivos ORC del mismo nivel que tienen el mismo nombre.

Por ejemplo, supongamos que desea asignar la tabla del ejemplo anterior, `SPECTRUM.ORC_EXAMPLE`, a un archivo ORC que utiliza la siguiente estructura de archivos.

```
• ORC file root(id = 0)
   o 'nested_col' : struct (id = 1)
      - 'map_col' : map (id = 2)
         - key : int (id = 3)
```

```
        - value : array (id = 4)
           - value : float (id = 5)
     - 'int_col' : int (id = 6)
  o 'int_col' : int (id = 7)
  o 'float_col' : float (id = 8)
```

Con la asignación por posición, Redshift Spectrum intentaría realizar la siguiente asignación.

| Nombre de columna de la tabla externa | ID de columna ORC | Nombre de columna ORC |
|---|---|---|
| int_col | 1 | struct |
| float_col | 7 | int_col |
| nested_col | 8 | float_col |

Si consulta una tabla con la asignación por posición anterior, el comando SELECT no realiza correctamente la validación de tipos porque las estructuras son diferentes.

Puede asignar la misma tabla externa a las dos estructuras de archivos que se muestran en los ejemplos anteriores utilizando la asignación por nombre de columna. Las columnas `int_col`, `float_col` y `nested_col` de la tabla se asignan por nombre a las columnas del archivo ORC que tienen el mismo nombre. La columna `nested_col` se la tabla externa es una columna `struct` con las subcolumnas `map_col` e `int_col`. Las subcolumnas también se asignan correctamente a las columnas correspondientes del archivo ORC a través del nombre.

## Creación de tablas externas para datos administrados en Apache Hudi

Si desea realizar consultas de datos en formato de copia en escritura (Copy On Write, CoW) de Apache Hudi, puede emplear las tablas externas de Amazon Redshift Spectrum. Una tabla de copia en escritura de Hudi consiste en una colección de archivos Apache Parquet almacenados en Amazon S3. Puede leer tablas de copia al escribir (CoW) en las versiones 0.5.2, 0.6.0, 0.7.0, 0.8.0, 0.9.0, 0.10.0, 0.10.1, 0.11.0 y 0,11.1 de Apache Hudi que se crean y modifican con las operaciones de escritura insert, delete y upsert. Por ejemplo, no se admiten tablas de arranque. Para obtener más información, consulte Tabla de copia en escritura en la documentación de Apache Hudi de código abierto.

Cuando se crea una tabla externa que referencia datos en formato CoW de Hudi, se debe asignar cada columna de la tabla externa a una columna de los datos Hudi. La asignación se realiza por columna.

Las instrucciones del lenguaje de definición de datos (DDL) para las tablas de Hudi con o sin partición se asemejan a las correspondientes a otros formatos de archivo Apache Parquet. En el caso de las tablas de Hudi, se define `INPUTFORMAT` como `org.apache.hudi.hadoop.HoodieParquetInputFormat`. El parámetro `LOCATION` debe apuntar a la carpeta base de la tabla de Hudi que contiene la carpeta `.hoodie`, la cual es necesaria para establecer la línea de tiempo de confirmación de Hudi. En algunos casos, una operación SELECT en una tabla de Hudi podría presentar un error y mostrar el mensaje No valid Hudi commit timeline found (No se encontró ninguna línea de tiempo de confirmación de Hudi válida). En ese caso, verifique si la carpeta `.hoodie` se encuentra en la ubicación correcta y contiene una línea de tiempo de confirmación de Hudi válida.

> **Note**
>
> El formato de Apache Hudi solo se admite cuando se utiliza un AWS Glue Data Catalog. No se admite cuando se utiliza un metastore de Apache Hive como el catálogo externo.

El DDL que permite definir una tabla sin partición presenta el siguiente formato.

```
CREATE EXTERNAL TABLE tbl_name (columns)
ROW FORMAT SERDE 'org.apache.hadoop.hive.ql.io.parquet.serde.ParquetHiveSerDe'
STORED AS
INPUTFORMAT 'org.apache.hudi.hadoop.HoodieParquetInputFormat'
OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat'
LOCATION 's3://s3-bucket/prefix'
```

El DDL que permite definir una tabla con partición presenta el siguiente formato.

```
CREATE EXTERNAL TABLE tbl_name (columns)
PARTITIONED BY(pcolumn1 pcolumn1-type[,...])
ROW FORMAT SERDE 'org.apache.hadoop.hive.ql.io.parquet.serde.ParquetHiveSerDe'
STORED AS
INPUTFORMAT 'org.apache.hudi.hadoop.HoodieParquetInputFormat'
OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat'
LOCATION 's3://s3-bucket/prefix'
```

Para agregar particiones a una tabla de Hudi con partición, ejecute un comando ALTER TABLE ADD PARTITION en el que el parámetro `LOCATION` apunte a la subcarpeta de Amazon S3 que contiene los archivos que pertenecen a la partición.

El DDL que permite agregar particiones presenta el siguiente formato.

```
ALTER TABLE tbl_name
ADD IF NOT EXISTS PARTITION(pcolumn1=pvalue1[,...])
LOCATION 's3://s3-bucket/prefix/partition-path'
```

## Creación de tablas externas para datos administrados en Delta Lake

Si desea realizar consultas sobre los datos de las tablas de Delta Lake, puede utilizar las tablas externas de Amazon Redshift Spectrum.

Si desea obtener acceso a una tabla de Delta Lake desde Redshift Spectrum, debe generar un manifiesto antes de realizar la consulta. Un manifiesto de Delta Lake contiene un listado de los archivos que conforman una instantánea consistente de la tabla de Delta Lake. En una tabla con partición, hay un manifiesto por cada partición. Una tabla de Delta Lake consiste en una colección de archivos Apache Parquet almacenados en Amazon S3. Para obtener más información, consulte Delta Lake en la documentación de Delta Lake de código abierto.

Cuando se crea una tabla externa que referencia los datos de las tablas de Delta Lake, se asigna cada columna de la tabla externa a una de las columnas de la tabla de Delta Lake. La asignación se realiza por nombre de columna.

El DDL para las tablas con o sin partición de Delta Lake se asemeja al correspondiente a otros formatos de archivo Apache Parquet. En el caso de las tablas de Delta Lake, define `INPUTFORMAT` como `org.apache.hadoop.hive.ql.io.SymlinkTextInputFormat` y `OUTPUTFORMAT` como `org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat`. El parámetro `LOCATION` debe apuntar a la carpeta del manifiesto en la carpeta base de la tabla. Si se produce un error en una operación SELECT de una tabla de Delta Lake, consulte Limitaciones y solución de problemas de las tablas de Delta Lake para obtener información sobre los posibles motivos.

El DDL que permite definir una tabla sin partición presenta el siguiente formato.

```
CREATE EXTERNAL TABLE tbl_name (columns)
ROW FORMAT SERDE 'org.apache.hadoop.hive.ql.io.parquet.serde.ParquetHiveSerDe'
STORED AS
```

```
INPUTFORMAT 'org.apache.hadoop.hive.ql.io.SymlinkTextInputFormat'
OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'
LOCATION 's3://s3-bucket/prefix/_symlink_format_manifest'
```

El DDL que permite definir una tabla con partición presenta el siguiente formato.

```
CREATE EXTERNAL TABLE tbl_name (columns)
PARTITIONED BY(pcolumn1 pcolumn1-type[,...])
ROW FORMAT SERDE 'org.apache.hadoop.hive.ql.io.parquet.serde.ParquetHiveSerDe'
STORED AS
INPUTFORMAT 'org.apache.hadoop.hive.ql.io.SymlinkTextInputFormat'
OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'
LOCATION 's3://s3-bucket>/prefix/_symlink_format_manifest'
```

Si desea agregar particiones a una tabla de Delta Lake con partición, ejecute un comando ALTER TABLE ADD PARTITION en el que el parámetro LOCATION apunte a la subcarpeta de Amazon S3 que contiene el manifiesto de la partición.

El DDL que permite agregar particiones presenta el siguiente formato.

```
ALTER TABLE tbl_name
ADD IF NOT EXISTS PARTITION(pcolumn1=pvalue1[,...])
LOCATION
's3://s3-bucket/prefix/_symlink_format_manifest/partition-path'
```

También puede ejecutar un DDL que apunte directamente al archivo de manifiesto de Delta Lake.

```
ALTER TABLE tbl_name
ADD IF NOT EXISTS PARTITION(pcolumn1=pvalue1[,...])
LOCATION
's3://s3-bucket/prefix/_symlink_format_manifest/partition-path/manifest'
```

## Limitaciones y solución de problemas de las tablas de Delta Lake

Cuando consulte las tablas de Delta Lake desde Redshift Spectrum, tenga en cuenta lo siguiente:

- Si un manifiesto apunta a una instantánea o una partición que ya no existe, se producen errores en las consultas hasta que se genere un nuevo manifiesto válido. Por ejemplo, esto podría suceder con una operación VACUUM en la tabla subyacente.
- Los manifiestos de Delta Lake solo proporcionan consistencia en el nivel de la partición.

En la siguiente tabla, se explican algunos de los posibles motivos por los cuales se producen ciertos errores cuando se consulta una tabla de Delta Lake.

| Mensaje de error | Posible motivo |
| --- | --- |
| El manifiesto de Delta Lake en el bucket s3-bucket-1 no puede contener entradas en el bucket s3-bucket-2. | Las entradas del manifiesto apuntan a archivos en un bucket de Amazon S3 distinto al especificado. |
| Se espera que los archivos de Delta Lake se encuentren en la misma carpeta. | Las entradas del manifiesto apuntan a archivos que poseen un prefijo de Amazon S3 distinto al especificado. |
| No se ha encontrado el archivo nombre del archivo que aparece en el manifiesto ruta de manifiesto de Delta Lake. | No se encontró en Amazon S3 un archivo incluido en el manifiesto. |
| Se produjo un error en la recuperación del manifiesto de Delta Lake. | No se encontró el manifiesto en Amazon S3. |
| La ruta de S3 no es válida. | Una entrada en el archivo de manifiesto no constituye una ruta válida de Amazon S3, o bien, se ha dañado el archivo de manifiesto. |

# Uso de tablas de Apache Iceberg con Amazon Redshift

En este tema se describe cómo utilizar las tablas en formato de Apache Iceberg con Redshift Spectrum o Redshift sin servidor. Apache Iceberg es un formato de alto rendimiento para tablas de análisis de gran tamaño.

Puede utilizar Redshift Spectrum o Redshift sin servidor para consultar las tablas de Apache Iceberg catalogadas en AWS Glue Data Catalog. Apache Iceberg es un formato de tabla de código abierto para lagos de datos. Para obtener más información, consulte Apache Iceberg en la documentación de Apache Iceberg.

Amazon Redshift proporciona coherencia transaccional para consultar las tablas de Apache Iceberg. Puede manipular los datos de las tablas mediante servicios compatibles con ACID (atomicidad, coherencia, aislamiento y durabilidad) como Amazon Athena y Amazon EMR mientras

ejecuta consultas con Amazon Redshift. Amazon Redshift puede utilizar las estadísticas de tabla almacenadas en los metadatos de Apache Iceberg para optimizar los planes de consulta y reducir los escaneos de archivos durante el procesamiento de las consultas. Con Amazon Redshift SQL, puede unir tablas de Redshift con tablas de lago de datos.

Para empezar a utilizar las tablas de Iceberg con Amazon Redshift:

1. Cree una tabla de Apache Iceberg en una base de datos de AWS Glue Data Catalog mediante un servicio compatible como Amazon Athena o Amazon EMR. Para crear una tabla de Iceberg mediante Athena, consulte Uso de las tablas de Apache Iceberg en la Guía del usuario de Amazon Athena.
2. Cree un clúster de Amazon Redshift o un grupo de trabajo de Redshift sin servidor con un rol de IAM asociado que permita el acceso a su lago de datos. Para obtener información sobre cómo crear clústeres o grupos de trabajo, consulte Introducción a los almacenamientos de datos de Amazon Redshift e Introducción a almacenamientos de datos de Redshift sin servidor en la Guía de introducción de Amazon Redshift.
3. Conéctese a su clúster o grupo de trabajo mediante el editor de consultas v2 o un cliente SQL de terceros. Para obtener información sobre cómo conectarse mediante el editor de consultas v2, consulte Conexión a un almacenamiento de datos de Amazon Redshift con herramientas de cliente de SQL en la Guía de administración de Amazon Redshift.
4. Cree un esquema externo en la base de datos de Amazon Redshift para una base de datos de catálogo de datos específica que incluya las tablas de Iceberg. Para obtener información sobre la creación de un esquema externo, consulte Esquemas externos en Amazon Redshift Spectrum.
5. Utilice las consultas SQL para acceder a las tablas Iceberg del esquema externo que ha creado.

## Consideraciones al utilizar tablas de Apache Iceberg con Amazon Redshift

Tenga en cuenta lo siguiente cuando utilice Amazon Redshift con tablas de Iceberg:

- Compatibilidad con la versión de Iceberg: Amazon Redshift admite la ejecución de consultas en las siguientes versiones de tablas de Iceberg:
  - Versión 1, que define cómo se administran las tablas analíticas de gran tamaño mediante archivos de datos inmutables.
  - Versión 2, que agrega la capacidad de admitir la actualización y eliminación en el nivel de fila sin modificar los archivos de datos existentes y el control de los cambios en los datos de la tabla mediante archivos de eliminación.

Para ver la diferencia entre las tablas de la versión 1 y la versión 2, consulte Format version changes en la documentación de Apache Iceberg.

- Solo consultas: Amazon Redshift admite el acceso de solo lectura a las tablas de Apache Iceberg. Admite consultas de selección transaccionales coherentes. Puede utilizar un servicio como Amazon Athena para definir y actualizar el esquema de las tablas de Iceberg en el AWS Glue Data Catalog.
- Adición de particiones: no es necesario que agregue manualmente particiones para sus tablas de Apache Iceberg. Amazon Redshift detecta automáticamente las nuevas particiones en las tablas de Apache Iceberg y no es necesaria ninguna operación manual para actualizar las particiones en la definición de tabla. Cualquier cambio en la especificación de las particiones también se aplica automáticamente a las consultas sin intervención del usuario.
- Ingesta de datos de Iceberg en Amazon Redshift: puede utilizar los comandos INSERT INTO o CREATE TABLE AS para importar datos de su tabla de Iceberg en una tabla local de Amazon Redshift. Actualmente no puede utilizar el comando COPY para ingerir el contenido de una tabla de Apache Iceberg en una tabla local de Amazon Redshift.
- Vistas materializadas: puede crear vistas materializadas en las tablas de Apache Iceberg como en cualquier otra tabla externa de Amazon Redshift. Se aplican a las tablas de Apache Iceberg las mismas consideraciones para otros formatos de tablas de lago de datos. Actualmente no se admiten las renovaciones automáticas, la reescritura automática de consultas ni las máquinas virtuales automáticas en las tablas del lago de datos.
- Control de acceso detallado de AWS Lake Formation Amazon Redshift admite el control de acceso detallado en las tablas de AWS Lake Formation.
- Parámetros de control de datos definidos por el usuario: Amazon Redshift admite parámetros de control de datos definidos por el usuario en tablas de Apache Iceberg. Utiliza parámetros de tratamiento de datos definidos por el usuario en los archivos existentes para adaptar los datos que se consultan en las tablas externas y evitar errores de análisis. Estos parámetros proporcionan capacidades para controlar las discordancias entre el esquema de la tabla y los datos reales de los archivos. También puede utilizar parámetros de control de datos definidos por el usuario en las tablas de Apache Iceberg.
- Consultas de viajes en el tiempo: las consultas de viajes en el tiempo no son compatibles actualmente con las tablas de Apache Iceberg.
- Precios: cuando acceda a las tablas de Iceberg desde un clúster, se le cobrarán los precios de Redshift Spectrum. Cuando acceda a las tablas de Iceberg desde un grupo de trabajo, se le

aplicarán los precios de Redshift sin servidor. Para obtener información sobre los precios de Redshift Spectrum y Redshift sin servidor, consulte Precios de Amazon Redshift.

## Tipos de datos compatibles con las tablas de Apache Iceberg

En este tema se describen los tipos de datos compatibles que Redshift Spectrum puede leer de tablas en formato de Apache Iceberg.

Amazon Redshift puede consultar las tablas de Iceberg que contienen los siguientes tipos de datos:

```
binary
boolean
date
decimal
double
float
int
list
long
map
string
struct
timestamp without time zone
```

Para obtener más información sobre los tipos de datos de Iceberg, consulte los esquemas para Iceberg en la documentación de Apache.

En la siguiente tabla, se muestra la relación entre los tipos de datos de Amazon Redshift y los tipos de datos de tabla de Iceberg.

| Tipo Iceberg | Tipo de Amazon Redshift | Notas |
|---|---|---|
| `boolean` | `boolean` | |
| - | `tinyint` | No compatible con tablas de Iceberg en Amazon Redshift. |
| - | `smallint` | No compatible con tablas de Iceberg en Amazon Redshift. |

| Tipo Iceberg | Tipo de Amazon Redshift | Notas |
|---|---|---|
| `int` | `int` | En las instrucciones SQL de Amazon Redshift, este tipo es `INTEGER`. |
| `long` | `bigint` | |
| `double` | `double precision` | |
| `float` | `real` | |
| `decimal(P, S)` | `decimal(P, S)` | `P` es la precisión, `S` es la escala. |
| - | `char` | No compatible con tablas de Iceberg en Redshift Spectrum. |
| `string` | `string` | En las instrucciones SQL de Amazon Redshift, este tipo es `VARCHAR`. |
| `binary` | `binary` | |
| `date` | `date` | |
| `time` | - | |
| `timestamp` | `timestamp` | |
| `timestamptz` | - | |
| `list<E>` | `array` | |
| `map<K,V>` | `map` | |

| Tipo Iceberg | Tipo de Amazon Redshift | Notas |
|---|---|---|
| `struct<...>` | `struct` | |
| `fixed(L)` | - | El tipo `fixed(L)` no se admite actualmente en Redshift Spectrum. |

Para obtener más información sobre los tipos de datos en Amazon Redshift, consulte Tipos de datos.

# Rendimiento de las consultas de Amazon Redshift Spectrum

En este tema se describe cómo mejorar el rendimiento de las consultas de Redshift Spectrum.

Observe el plan de consulta para saber qué pasos se han enviado a la capa de Amazon Redshift Spectrum.

Los siguientes pasos se relacionan con las consultas en Redshift Spectrum:

- S3 Seq Scan
- S3 HashAggregate
- S3 Query Scan
- Seq Scan PartitionInfo
- Partition Loop

En el siguiente ejemplo, se muestra el plan de consulta para una consulta que combina una tabla externa con una tabla local. Tenga en cuenta los pasos S3 Seq Scan y S3 HashAggregate que se ejecutaron sobre los datos en Amazon S3.

```
explain
select top 10 spectrum.sales.eventid, sum(spectrum.sales.pricepaid)
from spectrum.sales, event
where spectrum.sales.eventid = event.eventid
and spectrum.sales.pricepaid > 30
group by spectrum.sales.eventid
```

```
order by 2 desc;
```

```
QUERY PLAN

------------------------------------------------------------------------------
XN Limit  (cost=1001055770628.63..1001055770628.65 rows=10 width=31)

  ->  XN Merge  (cost=1001055770628.63..1001055770629.13 rows=200 width=31)

        Merge Key: sum(sales.derived_col2)

        ->  XN Network  (cost=1001055770628.63..1001055770629.13 rows=200 width=31)

              Send to leader

              ->  XN Sort  (cost=1001055770628.63..1001055770629.13 rows=200 width=31)

                    Sort Key: sum(sales.derived_col2)

                    ->  XN HashAggregate  (cost=1055770620.49..1055770620.99 rows=200
 width=31)

                          ->  XN Hash Join DS_BCAST_INNER  (cost=3119.97..1055769620.49
 rows=200000 width=31)

                                Hash Cond: ("outer".derived_col1 = "inner".eventid)

                                ->  XN S3 Query Scan sales  (cost=3010.00..5010.50
 rows=200000 width=31)

                                      ->  S3 HashAggregate  (cost=3010.00..3010.50
 rows=200000 width=16)
```

```
                                          ->  S3 Seq Scan spectrum.sales
location:"s3://redshift-downloads/tickit/spectrum/sales" format:TEXT
(cost=0.00..2150.00 rows=172000 width=16)
                                                Filter: (pricepaid > 30.00)


                               ->  XN Hash  (cost=87.98..87.98 rows=8798 width=4)


                                     ->  XN Seq Scan on event  (cost=0.00..87.98
rows=8798 width=4)
```

Tenga en cuenta los siguientes elementos en el plan de consulta:

- El nodo `S3 Seq Scan` muestra que el filtro `pricepaid > 30.00` se procesó en la capa de Redshift Spectrum.

  Un nodo de filtro bajo el nodo `XN S3 Query Scan` indica el procesamiento de predicados en Amazon Redshift sobre los datos devueltos desde la capa de Redshift Spectrum.
- El nodo `S3 HashAggregate` indica agregación en la capa de Redshift Spectrum para el grupo por cláusula (`group by spectrum.sales.eventid`).

A continuación, se muestran maneras de mejorar el rendimiento de Redshift Spectrum:

- Utilice archivos de datos con formato Apache Parquet. Parquet almacena datos en un formato de columnas, para que Redshift Spectrum pueda eliminar las columnas no necesarias para el examen. Cuando los datos están en formato de archivo de texto, Redshift Spectrum necesita examinar el archivo completo.
- Utilice distintos archivos para optimizar el procesamiento en paralelo. Asegúrese de que los archivos tienen un tamaño superior a 64 MB. Evite sesgos de tamaños de datos al mantener todos los archivos con, aproximadamente, el mismo tamaño. Para obtener información sobre los archivos de Apache Parquet y las recomendaciones de configuración, consulte File Format: Configurations en la Documentación de Apache Parquet.
- Utilice la menor cantidad posible de columnas en las consultas.
- Guarde las tablas de hechos grandes en Amazon S3 y mantenga las tablas de dimensiones más pequeñas y de uso frecuente en la base de datos local de Amazon Redshift.
- Actualice las estadísticas de la tabla externa al configurar el parámetro numRows del comando TABLE PROPERTIES. Utilice CREATE EXTERNAL TABLE o ALTER TABLE para establecer el

parámetro numRows de TABLE PROPERTIES de manera que refleje la cantidad de filas en la tabla. Amazon Redshift no analiza las tablas externas para generar las estadísticas de las tablas que el optimizador de consultas emplea a la hora de crear un plan de consulta. Si no se establecen las estadísticas de la tabla que corresponden a una tabla externa, Amazon Redshift crea un plan de ejecución de consultas. Amazon Redshift genera este plan a partir de la suposición de que las tablas externas son las más grandes, mientras que las tablas locales son las más pequeñas.

- El planificador de consultas de Amazon Redshift envía predicados y agrupaciones a la capa de consultas de Redshift Spectrum siempre que sea posible. Cuando se devuelven grandes cantidades de datos de Amazon S3, el procesamiento se ve limitado por los recursos del clúster. Redshift Spectrum escala automáticamente para procesar solicitudes grandes. Por lo tanto, el rendimiento general mejora siempre que pueda enviar el procesamiento a la capa de Redshift Spectrum.
- Escriba las consultas para utilizar filtros y agregaciones que sean legibles y que se puedan enviar a la capa de Redshift Spectrum.

  Los siguientes son ejemplos de algunas operaciones que se pueden enviar a la capa de Redshift Spectrum:

  - cláusulas GROUP BY
  - Condiciones de comparación y condiciones de coincidencia de patrones, como LIKE.
  - Funciones de agregación, como COUNT, SUM, AVG, MIN y MAX.
  - Funciones de cadena.

  Entre las operaciones que no se pueden enviar a la capa de Redshift Spectrum se incluyen DISTINCT y ORDER BY.
- Use particiones para limitar los datos por examinar. Particione los datos según los predicados más frecuentes de las consultas y, luego, depure las particiones al filtrar las columnas de partición. Para obtener más información, consulte Partición de tablas externas de Redshift Spectrum.

  Consulte SVL_S3PARTITION para ver las particiones totales y las particiones calificadas.
- Utilice el generador de estadísticas de AWS Glue para calcular las estadísticas en el nivel de columna para las tablas de AWS Glue Data Catalog. Una vez que AWS Glue genera las estadísticas para las tablas del catálogo de datos, Amazon Redshift Spectrum las utiliza automáticamente para optimizar el plan de consultas. Para obtener más información sobre la computación de estadísticas en el nivel de columna mediante AWS Glue, consulte Trabajar con las estadísticas de las columnas en la Guía para desarrolladores de AWS Glue.

# Opciones de gestión de datos

En este tema se describe cómo configurar la forma en que Redshift Spectrum gestiona los datos en formatos inesperados.

Puede establecer parámetros de tabla al crear tablas externas para adaptar los datos que se consultan en tablas externas. De lo contrario, se pueden producir errores de escaneo. Para obtener más información, consulte PROPIEDADES DE TABLA en CREATE EXTERNAL TABLE. Para ver ejemplos, consulta Ejemplos de control de datos. Para obtener una lista de errores, consulte SVL_SPECTRUM_SCAN_ERROR.

Puede establecer las siguientes PROPIEDADES DE TABLA al crear tablas externas para especificar el control de entradas de los datos que se consultan en tablas externas.

- `column_count_mismatch_handling` para identificar si el archivo contiene menos o más valores para una fila que el número de columnas especificado en la definición de tabla externa.
- `invalid_char_handling` para especificar el control de entrada de caracteres no válidos en columnas que contengan datos de VARCHAR, CHAR y cadenas. Cuando especifica REEMPLAZAR para `invalid_char_handling`, puede especificar el carácter de sustitución que se va a utilizar.
- `numeric_overflow_handling` para especificar el control de desbordamiento de conversión en columnas que contengan datos enteros y decimales.
- `surplus_bytes_handling` para especificar la gestión de entrada de los bytes sobrantes en las columnas que contienen datos VARBYTE.
- `surplus_char_handling` para especificar el control de entrada de caracteres sobrantes en columnas que contengan datos de VARCHAR, CHAR y cadenas.

Puede configurar una opción de configuración para cancelar consultas que superen un número máximo de errores. Para obtener más información, consulte spectrum_query_maxerror.

# Ejemplo: Realización de subconsultas correlacionadas en Redshift Spectrum

En este tema se describe cómo realizar subconsultas correlacionadas en Redshift Spectrum. Una subconsulta correlacionada es una consulta que utiliza valores de la consulta externa.

Puede realizar subconsultas correlacionadas en Redshift Spectrum. La pseudocolumna `$spectrum_oid` proporciona la capacidad de realizar consultas correlacionadas con Redshift Spectrum. Para realizar una subconsulta correlacionada, la pseudocolumna `$spectrum_oid` debe estar habilitada pero no aparece en la instrucción SQL. Para obtener más información, consulte Pseudocolumnas.

Para crear el esquema externo y las tablas externas de este ejemplo, consulte Introducción a Amazon Redshift Spectrum.

A continuación, se muestra un ejemplo de una subconsulta correlacionada en Redshift Spectrum.

```
select *
from myspectrum_schema.sales s
where exists
( select *
from myspectrum_schema.listing l
where l.listid = s.listid )
order by salesid
limit 5;
```

```
salesid  listid   sellerid   buyerid   eventid   dateid  qtysold  pricepaid
 commission   saletime
1        1        36861      21191     7872      1875    4        728         109.2
    2008-02-18 02:36:48
2        4        8117       11498     4337      1983    2        76          11.4
    2008-06-06 05:00:16
3        5        1616       17433     8647      1983    2        350         52.5
    2008-06-06 08:26:17
4        5        1616       19715     8647      1986    1        175         26.25
    2008-06-09 08:38:52
5        6        47402      14115     8240      2069    2        154         23.1
    2008-08-31 09:17:02
```

# Métricas en Amazon Redshift Spectrum

En este tema se describen las vistas del sistema que puede utilizar para supervisar consultas de Redshift Spectrum.

Puede supervisar las consultas de Amazon Redshift Spectrum con las siguientes vistas de sistema:

- SVL_S3QUERY

  Use la vista SVL_S3QUERY para obtener detalles sobre las consultas de Redshift Spectrum (consultas S3) a nivel segmento y sector del nodo.

- SVL_S3QUERY_SUMMARY

  Utilice la vista SVL_S3QUERY_SUMMARY para obtener un resumen de todas las consultas de Amazon Redshift Spectrum (consultas S3) que se ejecutaron en el sistema.

Los puntos a continuación son los elementos a tener en cuenta en SVL_S3QUERY_SUMMARY:

- La cantidad de filas que procesó la consulta de Redshift Spectrum.
- La cantidad de bytes analizados de Amazon S3. El costo de una consulta de Redshift Spectrum se ve reflejado en la cantidad de datos analizados de Amazon S3.
- La cantidad de bytes devueltos desde la capa Redshift Spectrum hasta el clúster. Una gran cantidad de datos devueltos podría afectar el rendimiento del sistema.
- La duración máxima y la duración promedio de las solicitudes de Redshift Spectrum. Si las solicitudes toman demasiado tiempo de ejecución, podría deberse a un cuello de botella.

# Solución de problemas de consultas en Amazon Redshift Spectrum

En este tema se describen problemas comunes que puede encontrar con las consultas de Amazon Redshift Spectrum.

Para ver los errores generados por las consultas de Redshift Spectrum, consulte la tabla de sistema SVL_S3LOG.

Temas

- Cantidad de reintentos superada
- Acceso limitado
- Límite de recursos superado
- No se han devuelto filas para una tabla particionada
- Error de no autorización
- Formatos de datos incompatibles

- Error de sintaxis al usar el DDL de Hive en Amazon Redshift
- Permiso para crear tablas temporales
- Rango no válido
- Número de versión de Parquet no válido

## Cantidad de reintentos superada

Si se agota el tiempo de espera de una solicitud de Amazon Redshift Spectrum, esta se cancela y se reenvía. Luego de cinco intentos fallidos, la consulta muestra el siguiente error.

```
error:  Spectrum Scan Error: Retries exceeded
```

Entre las causas posibles, se incluyen:

- Tamaños de archivo grandes (mayores que 1 GB). Verifique los tamaños de los archivos en Amazon S3 y busque archivos grandes y sesgos en el tamaño de los archivos. Divida los archivos grandes en archivos más pequeños, de entre 100 MB y 1 GB. Intente que todos los archivos tengan, aproximadamente, el mismo tamaño.
- Rendimiento lento de red. Ejecute la consulta en otro momento.

## Acceso limitado

Amazon Redshift Spectrum está sujeto a las Service Quotas de otros servicios de AWS. En caso de uso excesivo, es posible que las solicitudes de Redshift Spectrum deban procesarse más lento, lo que da como resultado el siguiente error.

```
error:  Spectrum Scan Error: Access throttled
```

Pueden ocurrir dos tipos de limitación:

- Acceso limitado por Amazon S3.
- Acceso limitado por AWS KMS.

El contexto de error proporciona más detalles sobre el tipo de limitación controlada. A continuación, puede encontrar causas y posibles resoluciones para esta limitación controlada.

## Acceso limitado por Amazon S3

Amazon S3 podría limitar una solicitud de Redshift Spectrum si la tasa de solicitudes de lectura en un prefijo es demasiado alta. Para obtener más información sobre una tasa de solicitudes GET /HEAD que puede alcanzar en Amazon S3, consulte Optimización del rendimiento de Amazon S3 en la Guía del usuario de Amazon Simple Storage Service. La tasa de solicitudes GET/HEAD de Amazon S3 tiene en cuenta todas las solicitudes GET/HEAD en un prefijo para que las diferentes aplicaciones que acceden al mismo prefijo compartan la tasa total de solicitudes.

Si Amazon S3 limita sus solicitudes de Redshift Spectrum con frecuencia, reduzca la cantidad de solicitudes GET/HEAD que Redshift Spectrum envía a Amazon S3. Para ello, intente combinar archivos pequeños en archivos más grandes. Le recomendamos utilizar archivos con un tamaño igual o superior a 64 MB.

Además, tenga en cuenta la posibilidad de particionar las tablas de Redshift Spectrum para aprovechar el filtrado temprano y reducir la cantidad de archivos a los que se obtiene acceso en Amazon S3. Para obtener más información, consulte Partición de tablas externas de Redshift Spectrum.

## Acceso limitado por AWS KMS

Si almacena los datos en Amazon S3 con cifrado del lado del servidor (SSE-S3 o SSE-KMS), Amazon S3 llama a una operación de API a AWS KMS para cada archivo al que acceda Redshift Spectrum. Estas solicitudes cuentan para la cuota de operaciones criptográficas; para obtener más información, consulte Solicitar cuotas de AWS KMS. Para obtener más información sobre SSE-S3 y SSE-KMS, consulte Protección de los datos con el cifrado del lado del servidor y Protección de los datos con el cifrado del lado del servidor mediante las claves KMS almacenadas en AWS KMS en la Guía del usuario de Amazon Simple Storage Service.

Un primer paso para reducir la cantidad de solicitudes que Redshift Spectrum realiza a AWS KMS consiste en reducir la cantidad de archivos a los que se accede. Para ello, intente combinar archivos pequeños en archivos más grandes. Le recomendamos utilizar archivos con un tamaño igual o superior a 64 MB.

Si AWS KMS limita sus solicitudes de Redshift Spectrum con frecuencia, considere la posibilidad de solicitar un aumento en la cuota de su tasa de solicitudes de AWS KMS para las operaciones criptográficas. Para solicitar un aumento de cuota, consulte Límites de los servicios de AWS en la Referencia general de Amazon Web Services.

## Límite de recursos superado

Redshift Spectrum aplica un límite superior sobre la cantidad de memoria que puede usar una solicitud. Una solicitud de Redshift Spectrum que requiere más memoria no logra ejecutarse, lo que produce el siguiente error.

```
error:  Spectrum Scan Error: Resource limit exceeded
```

Hay dos razones comunes que pueden hacer que una solicitud de Redshift Spectrum exceda su límite de memoria:

- Redshift Spectrum procesa una gran parte de los datos que no se puede dividir en fragmentos más pequeños.
- Redshift Spectrum procesa un paso de agrupación grande.

Recomendamos utilizar un formato de archivo que admita lecturas paralelas con tamaños divididos de 128 MB o menos. Consulte Archivos de datos para consultas en Amazon Redshift Spectrum para obtener información sobre formatos de archivo compatibles y directrices genéricas para la creación de archivos de datos. Cuando se utilizan formatos de archivo o algoritmos de compresión que no admiten lecturas paralelas, se recomienda mantener los tamaños de archivo entre 64 MB y 128 MB.

## No se han devuelto filas para una tabla particionada

Si la consulta devuelve cero filas de una tabla externa particionada, controle si se agregó una partición para esta tabla externa. Redshift Spectrum solo analiza los archivos en una ubicación de Amazon S3 que se haya agregado explícitamente con `ALTER TABLE … ADD PARTITION`. Consulte la vista SVV_EXTERNAL_PARTITIONS para encontrar las particiones existentes. Ejecute `ALTER TABLE … ADD PARTITION` para cada partición que falte.

## Error de no autorización

Verifique que el rol de IAM que corresponde al clúster permita acceder a los objetos del archivo de Amazon S3. Si la base de datos externa está en Amazon Athena, verifique que el rol de IAM permite el acceso a los recursos de Athena. Para obtener más información, consulte Políticas de IAM para Amazon Redshift Spectrum.

## Formatos de datos incompatibles

Para un formato de archivo en columnas, como Apache Parquet, el tipo de columna está integrado a los datos. El tipo de columna en la definición CREATE EXTERNAL TABLE debe coincidir con el tipo de columna en el archivo de datos. Si no coincide, recibirá un error similar al siguiente:

```
File 'https://s3bucket/location/file has an incompatible Parquet schema
for column ‘s3://s3bucket/location.col1'. Column type: VARCHAR, Par
```

El mensaje de error puede estar truncado debido a un límite en la longitud del mensaje. Para recuperar el mensaje de error completo, incluidos el nombre y el tipo de columna, consulte la vista de sistema SVL_S3LOG.

En el siguiente ejemplo, se consulta SVL_S3LOG para la última consulta realizada.

```
select message
from svl_s3log
where query = pg_last_query_id()
order by query,segment,slice;
```

A continuación, se muestra un ejemplo de un resultado que muestra el mensaje de error completo.

```
                            message
-------------------------------------------------------------------------------------------
Spectrum Scan Error. File 'https://s3bucket/location/file has an incompatible
Parquet schema for column ' s3bucket/location.col1'.
Column type: VARCHAR, Parquet schema:\noptional int64 l_orderkey [i:0 d:1 r:0]\n
```

Para corregir el error, altere la tabla externa para que coincida con el tipo de columna del archivo Parquet.

## Error de sintaxis al usar el DDL de Hive en Amazon Redshift

Amazon Redshift admite un lenguaje de definición de datos (DDL) para CREATE EXTERNAL TABLE que sea similar al DDL de Hive. No obstante, los dos tipos de DDL no siempre son exactamente iguales. Si copia el DDL de Hive para crear o modificar tablas externas de Amazon Redshift, es posible que enfrente errores de sintaxis. A continuación, se muestran ejemplos de diferencias entre Amazon Redshift y el DDL de Hive:

- Amazon Redshift requiere comillas simples ('), mientras que el DDL de Hive admite comillas dobles (").
- Amazon Redshift no admite el tipo de datos STRING. Utilice VARCHAR en su lugar.

## Permiso para crear tablas temporales

Para ejecutar consultas de Redshift Spectrum, el usuario de la base de datos debe tener permiso para crear tablas temporales en ella. En el siguiente ejemplo, se concede un permiso temporal para la base de datos `spectrumdb` al grupo de usuarios `spectrumusers`.

```
grant temp on database spectrumdb  to group spectrumusers;
```

Para obtener más información, consulte GRANT.

## Rango no válido

Redshift Spectrum espera que los archivos de Amazon S3 que pertenecen a una tabla externa no se sobrescriban durante una consulta. Si eso ocurre, puede provocar el siguiente error.

```
Error: HTTP response error code: 416 Message: InvalidRange The requested range is not
 satisfiable
```

Para evitar el error, asegúrese de que los archivos de Amazon S3 no se sobrescriban mientras se consultan con Redshift Spectrum.

## Número de versión de Parquet no válido

Redshift Spectrum comprueba los metadatos de cada archivo de Apache Parquet al que accede. Si la comprobación falla, puede provocar un error similar al siguiente:

```
File 'https://s3.region.amazonaws.com/s3bucket/location/file has an invalid version
 number
```

Existen dos razones habituales que pueden hacer que la comprobación falle:

- El archivo Parquet se ha sobrescrito durante la consulta (consulte Rango no válido ).
- El archivo de Parquet está dañado.

# Tutorial: Consulta de datos anidados con Amazon Redshift Spectrum

En este tutorial se muestra cómo consultar datos anidados con Redshift Spectrum. Los datos anidados son datos que contienen campos anidados. Los campos anidados son campos que se unen como una sola entidad, por ejemplo, matrices, estructuras u objetos.

Temas


- Descripción general
- Paso 1: Crear una tabla externa que contenga datos anidados
- Paso 2: Consultar los datos anidados en Amazon S3 con extensiones SQL
- Casos de uso de datos anidados
- Limitaciones de los datos anidados (versión preliminar)
- Serialización de datos JSON anidados y complejos


## Descripción general

Amazon Redshift Spectrum admite la consulta de datos anidados en los formatos de archivos Parquet, ORC, JSON e Ion. Redshift Spectrum obtiene acceso a los datos mediante el uso de tablas externas. Puede crear tablas externas que usen los tipos de datos complejos `struct`, `array` y `map`.

Supongamos, por ejemplo, que su archivo de datos contiene los siguientes datos en Amazon S3 en una carpeta llamada `customers`. Aunque no hay un solo elemento raíz, cada objeto JSON de estos datos de ejemplo representa una fila en una tabla.

```
{"id": 1,
 "name": {"given": "John", "family": "Smith"},
 "phones": ["123-457789"],
 "orders": [{"shipdate": "2018-03-01T11:59:59.000Z", "price": 100.50},
            {"shipdate": "2018-03-01T09:10:00.000Z", "price": 99.12}]
}
{"id": 2,
 "name": {"given": "Jenny", "family": "Doe"},
 "phones": ["858-8675309", "415-9876543"],
 "orders": []
}
{"id": 3,
```

```
 "name": {"given": "Andy", "family": "Jones"},
 "phones": [],
 "orders": [{"shipdate": "2018-03-02T08:02:15.000Z", "price": 13.50}]
}
```

Puede usar Amazon Redshift Spectrum para consultar datos anidados en archivos. El siguiente tutorial le muestra cómo hacerlo con los datos de Apache Parquet.

### Requisitos previos

Si aún no usa Redshift Spectrum, siga los pasos en el tutorial de Introducción a Amazon Redshift Spectrum antes de continuar.

Para crear un esquema externo, reemplace el ARN del rol de IAM en el siguiente comando por el ARN del rol que creó en Creación de un rol de IAM. Luego, ejecute el comando en cliente SQL.

```
create external schema spectrum
from data catalog
database 'myspectrum_db'
iam_role 'arn:aws:iam::123456789012:role/myspectrum_role'
create external database if not exists;
```

## Paso 1: Crear una tabla externa que contenga datos anidados

Puede ver el origen de datos descargándolo de Amazon S3.

Para crear la tabla externa para este tutorial, ejecute el siguiente comando.

```
CREATE EXTERNAL TABLE spectrum.customers (
  id     int,
  name   struct<given:varchar(20), family:varchar(20)>,
  phones array<varchar(20)>,
  orders array<struct<shipdate:timestamp, price:double precision>>
)
STORED AS PARQUET
LOCATION 's3://redshift-downloads/tickit/spectrum/customers/';
```

En el ejemplo anterior, la tabla externa `spectrum.customers` utiliza los tipos de datos `struct` y `array` para definir columnas con datos anidados. Amazon Redshift Spectrum admite la consulta de datos anidados en los formatos de archivos Parquet, ORC, JSON e Ion. El parámetro `STORED AS` es `PARQUET` para los archivos de Apache Parquet. El parámetro `LOCATION` tiene que hacer

referencia a la carpeta de Amazon S3 que contiene los datos o los archivos anidados. Para obtener más información, consulte CREATE EXTERNAL TABLE.

Puede anidar tipos `array` y `struct` en cualquier nivel. Por ejemplo, puede definir una columna denominada `toparray` tal como se muestra en el siguiente ejemplo.

```
toparray array<struct<nestedarray:
         array<struct<morenestedarray:
         array<string>>>>>
```

También puede anidar los tipos `struct` tal como se muestra para la columna `x` en el siguiente ejemplo.

```
x struct<a: string,
         b: struct<c: integer,
                   d: struct<e: string>
                  >
        >
```

## Paso 2: Consultar los datos anidados en Amazon S3 con extensiones SQL

Redshift Spectrum permite consultar los tipos complejos `array`, `map` y `struct` mediante extensiones de la sintaxis SQL de Amazon Redshift.

### Extensión 1: Acceso a columnas de estructuras

Puede extraer datos de columnas de `struct` mediante una notación de puntos que concatene nombres de campos para crear rutas. Por ejemplo, la siguiente consulta devuelve los nombres y los apellidos de los clientes. El acceso al nombre se consigue con la ruta larga `c.name.given`. El acceso al apellido se consigue con la ruta larga `c.name.family`.

```
SELECT c.id, c.name.given, c.name.family
FROM   spectrum.customers c;
```

La consulta anterior devuelve los siguientes datos.

```
id | given | family
---|-------|-------
1  | John  | Smith
2  | Jenny | Doe
```

```
3  | Andy  | Jones
(3 rows)
```

Una `struct` puede ser una columna de otra `struct`, que a su vez puede ser una columna de otra `struct`, en cualquier nivel. Las rutas que ofrecen acceso a columnas en `struct` profundamente anidadas pueden ser arbitrariamente largas. Por ejemplo, consulte la definición de la columna `x` a continuación.

```
x struct<a: string,
         b: struct<c: integer,
                   d: struct<e: string>
                  >
        >
```

Puede obtener acceso a los datos en `e` como `x.b.d.e`.

## Extensión 2: Desplazamiento por matrices en una cláusula FROM

Puede extraer datos de columnas `array` (y, por extensión, de columnas `map`) especificando las columnas `array` en una cláusula `FROM` en lugar de nombres de tablas. La extensión se aplica a la cláusula `FROM` de la consulta principal y también a las cláusulas `FROM` de las subconsultas.

Puede hacer referencia a los elementos `array` por posición como, por ejemplo, `c.orders[0]`. (versión preliminar)

Mediante la combinación del desplazamiento por `arrays` con las uniones, se pueden conseguir varios tipos de aplanamientos, tal como se explica en los siguientes casos de uso.

Aplanamiento mediante combinaciones interiores

La siguiente consulta selecciona los ID de cliente y las fechas de envío de los pedidos para los clientes que tienen pedidos. La extensión SQL en la cláusula FROM `c.orders o` depende del alias `c`.

```
SELECT c.id, o.shipdate
FROM   spectrum.customers c, c.orders o
```

Para cada cliente `c` que tiene pedidos, la cláusula `FROM` devuelve una fila para cada pedido `o` del cliente `c`. La fila combina la fila de cliente `c` y la fila de pedido `o`. A continuación, la cláusula `SELECT` solo conserva `c.id` y `o.shipdate`. El resultado es el siguiente.

```
id|      shipdate
--|----------------------
1 |2018-03-01  11:59:59
1 |2018-03-01  09:10:00
3 |2018-03-02  08:02:15
(3 rows)
```

El alias `c` proporciona acceso a los campos del cliente y el alias `o` proporciona acceso a los campos de los pedidos.

Las semánticas son similares al SQL estándar. Puede considerar una cláusula `FROM` como si ejecutara el siguiente bucle anidado, seguido de `SELECT` para elegir los campos de la salida.

```
for each customer c in spectrum.customers
  for each order o in c.orders
     output c.id and o.shipdate
```

Por consiguiente, si un cliente no tuviera un pedido, no aparecería en el resultado.

También puede considerar esto como la cláusula `FROM` que realiza una `JOIN` con la tabla `customers` y la matriz `orders`. De hecho, también puede escribir la consulta tal como se muestra en el siguiente ejemplo.

```
SELECT c.id, o.shipdate
FROM   spectrum.customers c INNER JOIN c.orders o ON true
```

> **Note**
>
> Si existe un esquema denominado `c` con una tabla llamada `orders`, entonces `c.orders` hará referencia a la tabla `orders` y no a la columna matriz de `customers`.

Aplanamiento mediante combinaciones izquierdas

La siguiente consulta devuelve todos los nombres de los clientes y sus pedidos. Si un cliente no ha realizado ningún pedido, también se devolverá el nombre del cliente. No obstante, en este caso, las columnas de pedido son NULL, tal y como se muestra en el siguiente ejemplo para Jenny Doe.

```
SELECT c.id, c.name.given, c.name.family, o.shipdate, o.price
```

```
FROM   spectrum.customers c LEFT JOIN c.orders o ON true
```

La consulta anterior devuelve los siguientes datos.

```
id  |  given  | family  |    shipdate          | price
----|---------|---------|----------------------|--------
 1  |  John   | Smith   | 2018-03-01  11:59:59 | 100.5
 1  |  John   | Smith   | 2018-03-01  09:10:00 |  99.12
 2  |  Jenny  | Doe     |                      |
 3  |  Andy   | Jones   | 2018-03-02  08:02:15 |  13.5
 (4 rows)
```

## Extensión 3: Acceso a una matriz de escalares directamente mediante un alias

Cuando un alias p en una cláusula FROM toma valores de una matriz de escalares, la consulta hace referencia a los valores de p como p. Por ejemplo, la siguiente consulta devuelve pares de nombres de clientes y números de teléfono.

```
SELECT c.name.given, c.name.family, p AS phone
FROM   spectrum.customers c LEFT JOIN c.phones p ON true
```

La consulta anterior devuelve los siguientes datos.

```
given  |  family  |  phone
-------|----------|-----------
John   | Smith    | 123-4577891
Jenny  | Doe      | 858-8675309
Jenny  | Doe      | 415-9876543
Andy   | Jones    |
(4 rows)
```

## Extensión 4: Acceso a elementos de asignaciones

Redshift Spectrum trata el tipo de dato `map` como un tipo `array` que contiene tipos `struct` con una columna `key` y una columna `value`. La columna `key` debe ser `scalar`; el valor puede ser cualquier tipo de dato.

Por ejemplo, el siguiente código crea una tabla externa con `map` para almacenar números de teléfono.

```
CREATE EXTERNAL TABLE spectrum.customers2 (
```

```
  id     int,
  name   struct<given:varchar(20), family:varchar(20)>,
  phones map<varchar(20), varchar(20)>,
  orders array<struct<shipdate:timestamp, price:double precision>>
)
STORED AS PARQUET
LOCATION 's3://redshift-downloads/tickit/spectrum/customers/';
```

Como un tipo `map` se comporta igual que un tipo `array` con columnas `key` y `value`, puede considerar los anteriores esquemas como si fueran el siguiente.

```
CREATE EXTERNAL TABLE spectrum.customers3 (
  id     int,
  name   struct<given:varchar(20), family:varchar(20)>,
  phones array<struct<key:varchar(20), value:varchar(20)>>,
  orders array<struct<shipdate:timestamp, price:double precision>>
)
STORED AS PARQUET
LOCATION 's3://redshift-downloads/tickit/spectrum/customers/';
```

La siguiente consulta devuelve los nombres de clientes con un número de teléfono móvil, y devuelve el número para cada nombre. La consulta de asignaciones se trata como si fuera el equivalente a consultar una `array` anidada de tipos `struct`. La siguiente consulta solo devolverá datos si creó la tabla externa tal y como se describió anteriormente.

```
SELECT c.name.given, c.name.family, p.value
FROM   spectrum.customers c, c.phones p
WHERE  p.key = 'mobile';
```

> **Note**
>
> La columna `key` para una `map` es una `string` para los tipos de archivo de Ion y JSON.

## Casos de uso de datos anidados

En este tema se describen casos de uso de datos anidados. Los datos anidados son datos que contienen campos anidados. Los campos anidados son campos que se unen como una sola entidad, por ejemplo, matrices, estructuras u objetos.

Puede combinar las extensiones descritas anteriormente con las características SQL habituales. Los siguientes casos de uso ilustran algunas combinaciones comunes. Estos ejemplos muestran cómo puede usar los datos anidados. No forman parte del tutorial.

Temas


- Obtención de datos anidados
- Agregación de datos anidados con subconsultas
- Combinación de datos de Amazon Redshift y datos anidados


## Obtención de datos anidados

Puede usar una instrucción `CREATE TABLE AS` para obtener datos de una tabla externa que contiene tipos de datos complejos. La siguiente consulta extrae todos los clientes y sus números de teléfono de la tabla externa mediante la instrucción `LEFT JOIN` y los almacena en la tabla `CustomerPhones` de Amazon Redshift.

```
CREATE TABLE CustomerPhones AS
SELECT  c.name.given, c.name.family, p AS phone
FROM    spectrum.customers c LEFT JOIN c.phones p ON true;
```

## Agregación de datos anidados con subconsultas

Puede usar una subconsulta para agregar datos anidados. El siguiente ejemplo ilustra este enfoque.

```
SELECT c.name.given, c.name.family, (SELECT COUNT(*) FROM c.orders o) AS ordercount
FROM   spectrum.customers c;
```

Se devuelven los siguientes datos.

```
given   |  family  |  ordercount
--------|----------|-------------
 Jenny  |  Doe     |       0
 John   |  Smith   |       2
 Andy   |  Jones   |       1
 (3 rows)
```

> **Note**
>
> Cuando agrega datos anidados agrupándolos por la fila principal, la forma más eficiente es la que se muestra en el ejemplo anterior. En dicho ejemplo, las filas anidadas de `c.orders` están agrupadas por su fila principal `c`. De forma alternativa, si sabe que el `id` es único para cada `customer` y `o.shipdate` nunca es nulo, puede realizar la agregación tal y como se muestra en el siguiente ejemplo. Sin embargo, este enfoque no es en general tan eficiente como el del ejemplo anterior.

```
SELECT    c.name.given, c.name.family, COUNT(o.shipdate) AS ordercount
FROM      spectrum.customers c LEFT JOIN c.orders o ON true
GROUP BY  c.id, c.name.given, c.name.family;
```

También puede escribir la consulta usando una subconsulta en la cláusula `FROM` que haga referencia a un alias (`c`) de la consulta antecesora y que extraiga datos de matriz. El siguiente ejemplo demuestra esta estrategia.

```
SELECT c.name.given, c.name.family, s.count AS ordercount
FROM   spectrum.customers c, (SELECT count(*) AS count FROM c.orders o) s;
```

## Combinación de datos de Amazon Redshift y datos anidados

También puede combinar datos de Amazon Redshift con datos anidados en una tabla externa. Por ejemplo, suponga que tiene los siguientes datos anidados en Amazon S3.

```
CREATE EXTERNAL TABLE spectrum.customers2 (
  id      int,
  name    struct<given:varchar(20), family:varchar(20)>,
  phones  array<varchar(20)>,
  orders  array<struct<shipdate:timestamp, item:int>>
);
```

Imagine también que tiene la siguiente tabla en Amazon Redshift.

```
CREATE TABLE prices (
  id int,
  price double precision
);
```

La siguiente consulta encuentra el número total y la cantidad de las compras de cada cliente en base a lo anterior. El siguiente ejemplo es solo una muestra. Solo devolverá datos si creó las tablas tal y como se describió anteriormente.

```
SELECT   c.name.given, c.name.family, COUNT(o.date) AS ordercount, SUM(p.price) AS
 ordersum
FROM     spectrum.customers2 c, c.orders o, prices p ON o.item = p.id
GROUP BY c.id, c.name.given, c.name.family;
```

## Limitaciones de los datos anidados (versión preliminar)

En este tema se describen las limitaciones de la lectura de datos anidados con Redshift Spectrum. Los datos anidados son datos que contienen campos anidados. Los campos anidados son campos que se unen como una sola entidad, por ejemplo, matrices, estructuras u objetos.

> **Note**
>
> Las limitaciones marcadas (versión preliminar) en la siguiente lista solo se aplican a la versión preliminar de los clústeres en las siguientes regiones.
>
> - Este de EE. UU. (Ohio) (us-east-2)
> - Este de EE. UU. (Norte de Virginia) (us-east-1)
> - EE. UU. Oeste (Norte de California) (us-west-1)
> - Asia-Pacífico (Tokio) (ap-northeast-1)
> - Europa (Irlanda) (eu-west-1)
> - Europa (Estocolmo) (eu-north-1)
>
> Para obtener información sobre la configuración de clústeres de versión preliminar, consulte Crear un clúster de versión preliminar en la Guía de administración de Amazon Redshift.

Las limitaciones siguientes son aplicables a los datos anidados:

- Un tipo `array` o `map` puede contener otros tipos `array` o `map` siempre y cuando las consultas que estén en las `arrays` o `maps` anidadas no devuelvan valores `scalar`. (versión preliminar)
- Amazon Redshift Spectrum solo admite tipos de datos complejos como tablas externas.
- Las columnas de resultados de las subconsultas deben ser de nivel superior. (versión preliminar)

- Si una expresión `OUTER JOIN` hace referencia a una tabla anidada, puede hacer referencia solo a esa tabla y a sus matrices anidadas (y asignaciones). Si una expresión `OUTER JOIN` no hace referencia a una tabla anidada, puede hacer referencia a cualquier número de tablas no anidadas.
- Si una cláusula `FROM` en una subconsulta hace referencia a una tabla anidada, no puede hacer referencia a ninguna otra tabla.
- Si una subconsulta depende de una tabla anidada que hace referencia a una tabla principal, la subconsulta solo puede usar la tabla principal en la cláusula `FROM`. No puede usar la principal en ninguna otra cláusula, como puede ser una cláusula `SELECT` o `WHERE`. Por ejemplo, la siguiente consulta no se ejecuta porque la `SELECT` cláusula de la subconsulta hace referencia a la tabla `c` principal.

  ```
  SELECT c.name.given
  FROM   spectrum.customers c
  WHERE (SELECT COUNT(c.id) FROM c.phones p WHERE p LIKE '858%') > 1;
  ```

  La siguiente consulta funciona porque el elemento principal `c` solo se usa en la cláusula `FROM` de la subconsulta.

  ```
  SELECT c.name.given
  FROM   spectrum.customers c
  WHERE (SELECT COUNT(*) FROM c.phones p WHERE p LIKE '858%') > 1;
  ```

- Una subconsulta que obtiene acceso a datos anidados desde cualquier lugar que no sea la cláusula `FROM` debe devolver un único valor. Las únicas excepciones son los operadores `(NOT) EXISTS` de las cláusulas `WHERE`.
- `(NOT) IN` no se admite.
- La profundidad máxima de anidamiento para todos los tipos anidados es de 100. Esta restricción se aplica a todos los formatos de archivo (Parquet, ORC, Ion y JSON).
- Las subconsultas de agregación que tienen acceso a los datos anidados solo pueden hacer referencia a `arrays` y `maps` en la cláusula `FROM`, y no a una tabla externa.
- No es posible realizar consultas de las pseudocolumnas de datos anidados en una tabla de Redshift Spectrum. Para obtener más información, consulte Pseudocolumnas.
- Al extraer datos de columnas de matrices o mapas al especificarlos en una cláusula `FROM`, solo puede seleccionar valores de esas columnas si los valores son `scalar`. Por ejemplo, las siguientes consultas intentan `SELECT` elementos de dentro de una matriz. La consulta que selecciona `arr.a` funciona porque `arr.a` es un valor `scalar`. La segunda consulta no funciona

porque `array` es una matriz extraída de `s3.nested table` de la cláusula `FROM`. (versión preliminar)

```
SELECT array_column FROM s3.nested_table;

array_column
-----------------
[{"a":1},{"b":2}]

SELECT arr.a FROM s3.nested_table t, t.array_column arr;

arr.a
-----
1

--This query fails to run.
SELECT array FROM s3.nested_table tab, tab.array_column array;
```

No puede utilizar una matriz o un mapa en la cláusula `FROM` que, a su vez, provenga de otra matriz o mapa. Para seleccionar matrices u otras estructuras complejas que estén anidadas dentro de otras matrices, considere la posibilidad de utilizar índices en la instrucción `SELECT`.

## Serialización de datos JSON anidados y complejos

En este tema se muestra cómo serializar datos anidados en formato JSON. Los datos anidados son datos que contienen campos anidados. Los campos anidados son campos que se unen como una sola entidad, por ejemplo, matrices, estructuras u objetos.

Una alternativa a los métodos demostrados en este tutorial es consultar columnas de recopilación anidada de nivel superior como datos JSON serializados. Puede utilizar la serialización para inspeccionar, convertir y capturar datos anidados como JSON con Redshift Spectrum. Este método es compatible con los formatos ORC, JSON, Ion y Parquet. Utilice el parámetro de configuración de sesión `json_serialization_enable` para establecer el comportamiento de serialización. Una vez que se establece, los tipos de datos JSON complejos se serializan en VARCHAR(65535). Se puede acceder a los datos JSON anidados con Funciones JSON. Para obtener más información, consulte json_serialization_enable.

Por ejemplo, sin establecer `json_serialization_enable`, las siguientes consultas que acceden directamente a columnas anidadas producen error.

```
SELECT * FROM spectrum.customers LIMIT 1;

=> ERROR:  Nested tables do not support '*' in the SELECT clause.

SELECT name FROM spectrum.customers LIMIT 1;

=> ERROR:  column "name" does not exist in customers
```

La configuración de `json_serialization_enable` permite consultar directamente las recopilaciones de nivel superior.

```
SET json_serialization_enable TO true;

SELECT * FROM spectrum.customers order by id LIMIT 1;

id | name                                  | phones         | orders
---+---------------------------------------+----------------
+-----------------------------------------------------------------------------------------------------------------
1  | {"given": "John", "family": "Smith"} | ["123-457789"] | [{"shipdate":
 "2018-03-01T11:59:59.000Z", "price": 100.50}, {"shipdate": "2018-03-01T09:10:00.000Z",
 "price": 99.12}]

SELECT name FROM spectrum.customers order by id LIMIT 1;

name
---------
{"given": "John", "family": "Smith"}
```

Tenga en cuenta los siguientes elementos a la hora de serializar los datos JSON anidados.

- Cuando las columnas de recopilación se serializan como VARCHAR(65535), no se puede acceder directamente a sus subcampos anidados como parte de la sintaxis de la consulta (por ejemplo, en la cláusula de filtro). Sin embargo, las funciones JSON se pueden usar para acceder a datos JSON anidados.
- No se admiten las siguientes representaciones especializadas:
  - uniones ORC
  - mapas ORC con claves de tipo complejo
  - datagramas Ion
  - Ion SEXP

- Las marcas temporales se devuelven como cadenas serializadas ISO.
- Las claves de mapa primitivas se promueven a la categoría de cadena (por ejemplo, de 1 a "1").
- Los valores nulos de nivel superior se serializan como NULL.
- Si la serialización supera el tamaño máximo de VARCHAR de 65535, la celda se establece como NULL.

## Serialización de tipos complejos que contienen cadenas JSON

De manera predeterminada, los valores de cadena contenidos en recopilaciones anidadas se serializan como cadenas JSON con escape. El escape puede ser no deseado cuando las cadenas son datos JSON válidos. En cambio, es posible que desee escribir directamente como JSON subelementos anidados o campos VARCHAR. Habilite este comportamiento con la configuración de `json_serialization_parse_nested_strings` a nivel de la sesión. Cuando se establecen `json_serialization_enable` y `json_serialization_parse_nested_strings`, los valores JSON válidos se serializan en línea sin caracteres de escape. Cuando el valor no tenga el formato JSON válido, se escapa como si el valor de configuración `json_serialization_parse_nested_strings` no se hubiera establecido. Para obtener más información, consulte json_serialization_parse_nested_strings.

Por ejemplo, suponga que los datos del ejemplo anterior contenían JSON como un tipo complejo de `structs` en el campo `name` VARCHAR(20):

```
name
---------
{"given": "{\"first\":\"John\",\"middle\":\"James\"}", "family": "Smith"}
```

Cuando se haya establecido `json_serialization_parse_nested_strings`, la columna `name` se serializa de la siguiente manera:

```
SET json_serialization_enable TO true;
SET json_serialization_parse_nested_strings TO true;
SELECT name FROM spectrum.customers order by id LIMIT 1;

name
---------
{"given": {"first":"John","middle":"James"}, "family": "Smith"}
```

En lugar del siguiente escape:

```
SET json_serialization_enable TO true;
SELECT name FROM spectrum.customers order by id LIMIT 1;

name
---------
{"given": "{\"first\":\"John\",\"middle\":\"James\"}", "family": "Smith"}
```

# Bocetos de HyperLogLog

En este tema se describe cómo utilizar los bocetos de HyperLogLog en Amazon Redshift. HyperlogLog es un algoritmo para el problema del recuento distinto, que aproxima el número de elementos distintos de un conjunto de datos. Los bocetos de HyperLogLog son matrices de datos exclusivos sobre un conjunto de datos.

HyperLogLog es un algoritmo que se utiliza para estimar la cardinalidad de un conjunto múltiple. Cardinalidad se refiere al número de valores distintos en un conjunto múltiple. Por ejemplo, en el conjunto de {4,3,6,2,2,6,4,3,6,2,2,3}, la cardinalidad es 4 con valores distintos de 4, 3, 6 y 2.

La precisión del algoritmo HyperLogLog (que también se conoce como valor m) puede afectar la precisión de la cardinalidad estimada. Durante la estimación de la cardinalidad, Amazon Redshift utiliza un valor de precisión predeterminado de 15. Este valor puede ser de hasta 26 para conjuntos de datos más pequeños. Por lo tanto, el error relativo promedio oscila entre 0,01 y 0,6 %.

Al calcular la cardinalidad de un conjunto múltiple, el algoritmo HyperLogLog genera una construcción denominada boceto HLL. Un boceto HLL encapsula información sobre los valores distintos en un conjunto múltiple. El tipo de datos de Amazon Redshift HLLSKETCH representa dichos valores de boceto. Este tipo de datos se puede utilizar para almacenar bocetos en una tabla de Amazon Redshift. Además, Amazon Redshift es compatible con operaciones que se pueden aplicar a valores HLLSKETCH como funciones de agrupación y escalares. Puede utilizar estas funciones para extraer la cardinalidad de un HLLSKETCH y combinar varios valores HLLSKETCH.

El tipo de datos HLLSKETCH ofrece importantes beneficios de rendimiento de la consulta al extraer la cardinalidad de conjuntos de datos grandes. Puede agrupar previamente estos conjuntos de datos utilizando valores HLLSKETCH y almacenarlos en tablas. Amazon Redshift puede extraer la cardinalidad de manera directa de los valores HLLSKETCH almacenados sin acceder a los conjuntos de datos subyacentes.

Al procesar bocetos HLL, Amazon Redshift realiza optimizaciones que minimizan la huella de memoria del boceto y maximizan la precisión de la cardinalidad extraída. Amazon Redshift utiliza dos representaciones para bocetos HLL, dispersos y densos. Un HLLSKETCH comienza en formato disperso. A medida que se insertan nuevos valores en él, su tamaño aumenta. Una vez que su tamaño alcance el de la representación densa, Amazon Redshift convierte automáticamente el boceto de disperso a denso.

Amazon Redshift importa, exporta e imprime un HLLSKETCH como JSON cuando el boceto tiene un formato disperso. Amazon Redshift importa, exporta e imprime un HLLSKETCH como cadena

Base64 cuando el boceto tiene un formato denso. Para obtener más información sobre UNLOAD, consulte Descarga del tipo de datos HLLSKETCH. Para importar datos de valores separados por comas (CSV) o texto en Amazon Redshift, utilice el comando COPY. Para obtener más información, consulte Carga del tipo de datos HLLSKETCH.

Para obtener información acerca de las funciones utilizadas con HyperLogLog, consulte Funciones HyperLogLog.

Temas


- Consideraciones
- Limitaciones
- Ejemplos


# Consideraciones

En este tema se describen los detalles de uso de HyperLogLog en Amazon Redshift.

A continuación, se detallan consideraciones para utilizar HyperLogLog en Amazon Redshift:

- Las siguientes funciones que no corresponden a HyperLogLog pueden aceptar una entrada de tipo HLLSKETCH o columnas de tipo HLLSKETCH:
  - La función de agregación COUNT
  - Las expresiones condicionales COALESCE y NVL
  - Expresiones CASE
- La codificación compatible es RAW.
- Puede realizar una operación UNLOAD en tabla con columnas HLLSKETCH en texto o CSV. Puede utilizar las columnas UNLOAD HLLSKETCH para escribir datos de HLLSKETCH. Amazon Redshift muestra los datos en un formato JSON para una representación dispersa o un formato Base64 para una representación densa. Para obtener más información sobre UNLOAD, consulte Descarga del tipo de datos HLLSKETCH.

  A continuación, se muestra el formato utilizado para un boceto de HyperLogLog disperso representado en formato JSON.

```
{"version":1,"logm":15,"sparse":{"indices":
[15099259,33107846,37891580,50065963],"values":[2,3,2,1]}}
```

- Puede importar texto o datos CSV a Amazon Redshift con el comando COPY. Para obtener más información, consulte Carga del tipo de datos HLLSKETCH.
- La codificación predeterminada para HLLSKETCH es RAW. Para obtener más información, consulte Codificaciones de compresión.

## Limitaciones

En este tema se describen las limitaciones de HyperLogLog en Amazon Redshift.

A continuación, se detallan limitaciones para utilizar HyperLogLog en Amazon Redshift:

- Las tablas de Amazon Redshift no son compatibles con una columna HLLSKETCH como clave de ordenación o clave de distribución para una tabla de Amazon Redshift.
- Amazon Redshift no es compatible con columnas HLLSKETCH en cláusulas ORDER BY, GROUP BY o DISTINCT.
- Solo puede descargar columnas HLLSKETCH en formato de texto o CSV. A continuación, Amazon Redshift escribe los datos de HLLSKETCH en formato JSON o Base64. Para obtener más información sobre UNLOAD, consulte UNLOAD.
- Amazon Redshift solo admite los bocetos de HyperLogLog con una precisión (valor logm) de 15.
- Los controladores JDBC y ODBC no son compatibles con el tipo de datos HLLSKETCH. Por lo tanto, el conjunto de resultados utiliza VARCHAR para representar los valores HLLSKETCH.
- Amazon Redshift Spectrum no es compatible de forma nativa con los datos HLLSKETCH. Por lo tanto, no puede crear ni cambiar una tabla externa con una columna HLLSKETCH.
- Los tipos de datos para funciones definidas por el usuario (UDF) de Python no admiten el tipo de datos HLLSKETCH. Para obtener más información acerca de las UDF de Python, consulte UDF de Python escalares.

## Ejemplos

En esta sección se incluyen ejemplos del uso de HyperLogLog con Amazon Redshift.

Temas

- Ejemplo: devolver la cardinalidad en una subconsulta
- Ejemplo: devolver un tipo HLLSKETCH a partir de bocetos combinados en una subconsulta
- Ejemplo: devolver un boceto de HyperLogLog a partir de la combinación de varios bocetos

- Ejemplo: generar bocetos de HyperLogLog a partir de datos de S3 mediante tablas externas

## Ejemplo: devolver la cardinalidad en una subconsulta

En el siguiente ejemplo, se devuelve la cardinalidad de cada boceto en una subconsulta para una tabla denominada Ventas.

```
CREATE TABLE Sales (customer VARCHAR, country VARCHAR, amount BIGINT);
INSERT INTO Sales VALUES ('David Joe', 'Greece', 14.5),  ('David Joe', 'Greece',
 19.95), ('John Doe', 'USA', 29.95), ('John Doe', 'USA', 19.95), ('George Spanos',
 'Greece', 9.95), ('George Spanos', 'Greece', 2.95);
```

La siguiente consulta genera un boceto HLL para los clientes de cada país y extrae la cardinalidad. Esto muestra clientes únicos de cada país.

```
SELECT hll_cardinality(sketch), country
FROM (SELECT hll_create_sketch(customer) AS sketch, country
        FROM Sales
        GROUP BY country) AS hll_subquery;

hll_cardinality | country
----------------+---------
            1   | USA
            2   | Greece
 ...
```

## Ejemplo: devolver un tipo HLLSKETCH a partir de bocetos combinados en una subconsulta

En el siguiente ejemplo, se devuelve un único tipo HLLSKETCH que representa la combinación de bocetos individuales a partir de una subconsulta. Los bocetos se combinan mediante la función de agrupación HLL_COMBINE.

```
SELECT hll_combine(sketch)
FROM (SELECT hll_create_sketch(customer) AS sketch
        FROM Sales
        GROUP BY country) AS hll_subquery

                                        hll_combine
```

```
----------------------------------------------------------------------------------------------
 {"version":1,"logm":15,"sparse":{"indices":[29808639,35021072,47612452],"values":
[1,1,1]}}
(1 row)
```

## Ejemplo: devolver un boceto de HyperLogLog a partir de la combinación de varios bocetos

Para el siguiente ejemplo, suponga que la tabla `page-users` almacena bocetos agrupados previamente para cada página que los usuarios visitaron en un sitio web determinado. Cada fila de esta tabla contiene un boceto de HyperLogLog que representa todos los ID de usuario que muestran las páginas visitadas.

```
page_users
-- +----------------+-------------+--------------+
-- | _PARTITIONTIME | page         | sketch |
-- +----------------+-------------+--------------+
-- | 2019-07-28     | homepage     | CHAQkAQYA... |
-- | 2019-07-28     | Product A    | CHAQxPnYB... |
-- +----------------+-------------+--------------+
```

En el siguiente ejemplo, se unen los bocetos agrupados previamente y se genera un solo boceto. Este boceto encapsula la cardinalidad colectiva que abarca cada boceto.

```
SELECT hll_combine(sketch) as sketch
FROM page_users
```

El resultado tiene un aspecto similar al siguiente.

```
-- +----------------------------------------+
-- | sketch |
-- +----------------------------------------+
-- | CHAQ3sGoCxgCIAuCB4iAIBgTIBgqgIAgAwY.... |
-- +----------------------------------------+
```

Cuando se crea un nuevo boceto, puede utilizar la función HLL_CARDINALITY para obtener los valores distintos colectivos, como se muestra a continuación.

```
SELECT hll_cardinality(sketch)
```

```
FROM (
  SELECT
  hll_combine(sketch) as sketch
  FROM page_users
) AS hll_subquery
```

El resultado tiene un aspecto similar al siguiente.

```
-- +-------+
-- | count |
-- +-------+
-- | 54356 |
-- +-------+
```

## Ejemplo: generar bocetos de HyperLogLog a partir de datos de S3 mediante tablas externas

Los siguientes ejemplos almacenan en caché los bocetos de HyperLogLog para evitar el acceso directo a Amazon S3 para la estimación de la cardinalidad.

Puede agrupar previamente y almacenar en caché bocetos de HyperLogLog en tablas externas definidas para contener datos de Amazon S3. Al hacer esto, puede extraer estimaciones de cardinalidad sin acceder a los datos base subyacentes.

Por ejemplo, suponga que ha llevado un conjunto de archivos de texto delimitado por tabulaciones en Amazon S3. Ejecuta la siguiente consulta para definir una tabla externa denominada `sales` en el esquema externo de Amazon Redshift denominado `spectrum`. El bucket de Amazon S3 para este ejemplo está en la Región de AWS Este de EE. UU. (Norte de Virginia).

```
create external table spectrum.sales(
salesid integer,
listid integer,
sellerid smallint,
buyerid smallint,
eventid integer,
dateid integer,
qtysold integer,
pricepaid decimal(8,2),
commission decimal(8,2),
saletime timestamp)
```

```
row format delimited
fields terminated by '\t' stored as textfile
location 's3://redshift-downloads/tickit/spectrum/sales/';
```

Supongamos que desea calcular la cantidad de compradores distintos que adquirieron un elemento en fechas arbitrarias. Para ello, en el siguiente ejemplo, se generan bocetos para los ID de comprador para cada día del año y se almacena el resultado en la tabla `hll_sales` de Amazon Redshift.

```
CREATE TABLE hll_sales AS
SELECT saletime, hll_create_sketch(buyerid) AS sketch
FROM spectrum.sales
GROUP BY saletime;

SELECT TOP 5 * FROM hll_sales;
```

El resultado tiene un aspecto similar al siguiente.

```
-- hll_sales

-- | saletime        | sketch
    |
-- +-----------------
+-----------------------------------------------------------------------+
-- | 7/22/2008 8:30  | {"version":1,"logm":15,"sparse":{"indices":[9281416],"values":
[1]}}
-- | 2/19/2008 0:38  | {"version":1,"logm":15,"sparse":{"indices":[48735497],"values":
[3]}}
-- | 11/5/2008 4:49  | {"version":1,"logm":15,"sparse":{"indices":[27858661],"values":
[1]}}
-- | 10/27/2008 4:08 | {"version":1,"logm":15,"sparse":{"indices":[65295430],"values":
[2]}}
-- | 2/16/2008 9:37  | {"version":1,"logm":15,"sparse":{"indices":[56869618],"values":
[2]}}
-- +----------------
 +-----------------------------------------------------------------------+
```

La siguiente consulta muestra el número estimado de compradores distintos que adquirieron un elemento durante el viernes después de Acción de Gracias en 2008.

```
SELECT hll_cardinality(hll_combine(sketch)) as distinct_buyers
```

```
FROM hll_sales
WHERE trunc(saletime) = '2008-11-28';
```

El resultado tiene un aspecto similar al siguiente.

```
distinct_buyers
---------------
386
```

Suponga que desea conocer el número de usuarios distintos que compraron un elemento en un cierto intervalo de fechas. Un ejemplo podría ser desde el viernes después de Acción de Gracias hasta el lunes siguiente. Para obtener esto, la siguiente consulta utiliza la función de agrupación `hll_combine`. Esta función le permite evitar el recuento doble de los compradores que adquirieron un elemento en más de un día del rango seleccionado.

```
SELECT hll_cardinality(hll_combine(sketch)) as distinct_buyers
FROM hll_sales
WHERE saletime BETWEEN '2008-11-28' AND '2008-12-01';
```

El resultado tiene un aspecto similar al siguiente.

```
distinct_buyers
---------------
1166
```

Para mantener la tabla `hll_sales` actualizada, ejecute la siguiente consulta al final de cada día. Al hacerlo, se genera un boceto de HyperLogLog en función de los ID de los compradores que adquirieron un elemento hoy y lo agregan a la tabla `hll_sales`.

```
INSERT INTO hll_sales
SELECT saletime, hll_create_sketch(buyerid)
FROM spectrum.sales
WHERE TRUNC(saletime) = to_char(GETDATE(), 'YYYY-MM-DD')
GROUP BY saletime;
```

# Consultas entre bases de datos

En este tema se describen las consultas entre bases de datos, que son consultas que funcionan en varias bases de datos de Amazon Redshift dentro de un único clúster de Amazon Redshift.

Mediante el uso de consultas entre bases de datos en Amazon Redshift, puede consultar diferentes bases de datos y escribir en ellas en un clúster de Amazon Redshift. Con las consultas entre bases de datos, puede consultar y escribir datos desde cualquier base de datos del clúster de Amazon Redshift, independientemente de la base de datos a la que esté conectado. Las consultas entre bases de datos eliminan las copias de datos y simplifican la organización de datos para admitir varios grupos empresariales del mismo almacenamiento de datos.

Gracias a las consultas entre bases de datos, puede hacer lo siguiente:

- Consulte datos entre bases de datos del clúster de Amazon Redshift.

  No solo puede consultar desde bases de datos a las que está conectado, sino que también puede leer desde cualquier otra base de datos para la que tenga permisos.

  Cuando consulta objetos de una base de datos en otras bases de datos desconectadas, solo tiene acceso de lectura a esos objetos de la base de datos. Puede utilizar las consultas entre bases de datos para acceder a los datos de cualquiera de las bases de datos del clúster de Amazon Redshift sin tener que conectarse a esa base de datos específica. Esto puede ayudarlo a consultar y unir datos distribuidos en varias bases de datos del clúster de Amazon Redshift de forma rápida y sencilla.

  También puede unir conjuntos de datos de varias bases de datos en una sola consulta y analizar los datos con herramientas de inteligencia empresarial (BI) o análisis. Puede seguir configurando controles de acceso de lectura de tabla pormenorizados para los usuarios mediante comandos de SQL estándar de Amazon Redshift. De este modo, ayuda a garantizar que los usuarios solo vean los subconjuntos relevantes de los datos para los que tienen permisos.
- Escriba datos en diversas bases de datos en su clúster de Amazon Redshift.

  Puede escribir desde las bases de datos a las que esté conectado y también desde cualquier otra base de datos para la que tenga permisos.

  Cuando dispone de permisos de escritura en objetos de base de datos en cualquier otra base de datos no conectada, puede utilizar consultas entre bases de datos para escribir datos desde cualquier base de datos en su clúster de Amazon Redshift sin tener que conectarse a esa base de

datos específica. Esto puede ayudarlo a realizar operaciones de escritura complejas a la vez que une de forma rápida y sencilla datos que están repartidos en varias bases de datos en su clúster de Amazon Redshift.

También puede unir conjuntos de datos de varias bases de datos en una sola consulta y escribir los datos utilizando diferentes herramientas de extracción, transformación y carga (ETL) o de análisis. Puede seguir configurando controles de acceso de escritura de tabla pormenorizados para los usuarios mediante comandos de SQL estándar de Amazon Redshift. Esto garantiza que los usuarios vean solo los subconjuntos adecuados de los datos para los que tienen permisos.

- Consulte objetos.

  Puede consultar otros objetos de base de datos mediante nombres de objetos completamente calificados expresados con la notación de tres partes. La ruta completa a cualquier objeto de base de datos consta de tres componentes: el nombre de base de datos, el esquema y el nombre del objeto. Puede acceder a cualquier objeto desde cualquier otra base de datos con la notación de ruta completa, *database_name*.*schema_name*.*object_name*. Para acceder a una columna en particular, utilice *database_name*.*schema_name*.*object_name*.*column_name*.

  También puede crear un alias para un esquema en otra base de datos mediante la notación de esquema externo. Este esquema externo referencia otro par de base de datos y esquema. La consulta puede acceder al otro objeto de la base de datos mediante la notación de esquema externo, *external_schema_name*.*object_name*.

  En la misma consulta de solo lectura, puede consultar varios objetos de base de datos, como tablas de usuario, vistas regulares, vistas materializadas y vistas de enlace de tiempo de ejecución de otras bases de datos.

- Administre permisos.

  Los usuarios con privilegios de acceso a los objetos de cualquier base de datos de un clúster de Amazon Redshift pueden consultar y escribir datos en dichos objetos. Concede privilegios a usuarios y grupos de usuarios mediante el comando GRANT. También puede revocar los privilegios con el comando REVOKE cuando un usuario ya no requiere acceso a objetos de base de datos específicos.

- Trabaje con metadatos y herramientas de BI.

  Puede crear un esquema externo para hacer referencia a un esquema en otra base de datos de Amazon Redshift dentro del mismo clúster de Amazon Redshift. Para obtener más información, consulte el comando CREATE EXTERNAL SCHEMA.

Una vez que se crean referencias del esquema externo, Amazon Redshift muestra las tablas en el esquema de la otra base de datos en SVV_EXTERNAL_TABLES y SVV_EXTERNAL_COLUMNS para que las herramientas exploren los metadatos.

Para integrar consultas entre bases de datos a herramientas de BI, puede utilizar las siguientes vistas del sistema. Esto lo ayuda a ver información acerca de los metadatos de los objetos de las bases de datos conectadas y de otras bases de datos del clúster de Amazon Redshift.

A continuación, se presentan vistas del sistema que muestran todos los objetos de Amazon Redshift y los objetos externos de todas las bases de datos del clúster de Amazon Redshift:

- SVV_ALL_COLUMNS
- SVV_ALL_SCHEMAS
- SVV_ALL_TABLES

A continuación, se presentan vistas del sistema que muestran todos los objetos de Amazon Redshift de todas las bases de datos del clúster de Amazon Redshift:

- SVV_REDSHIFT_COLUMNS
- SVV_REDSHIFT_DATABASES
- SVV_REDSHIFT_FUNCTIONS
- SVV_REDSHIFT_SCHEMAS
- SVV_REDSHIFT_TABLES

Temas


- Consideraciones
- Limitaciones
- Ejemplos de consultas entre bases de datos
- Uso de consultas entre bases de datos con el editor de consultas


# Consideraciones

Este tema describe los detalles de uso de las consultas entre bases de datos en Amazon Redshift.

Cuando trabaje con la característica de consulta entre bases de datos de Amazon Redshift, tenga en cuenta lo siguiente:

- Amazon Redshift admite consultas entre bases de datos en todos los tipos de nodos ra3 y espacios de nombres sin servidor.
- Amazon Redshift admite la unión de datos de tablas o vistas en una o más bases de datos del mismo clúster de Amazon Redshift.
- Todas las consultas de una transacción en la base de datos conectada leen datos en el mismo estado de la otra base de datos en que los datos estaban al principio de la transacción. Este enfoque ayuda a proporcionar consistencia transaccional de consultas entre bases de datos. Amazon Redshift admite la consistencia transaccional para las consultas entre bases de datos.
- Para obtener metadatos en las bases de datos, utilice las vistas de metadatos SVV_ALL* y SVV_REDSHIFT*. No puede utilizar la notación de tres partes ni los esquemas externos para consultar tablas o vistas de metadatos entre bases de datos en information_schema y pg_catalog.

# Limitaciones

Este tema describe limitaciones para consultas entre bases de datos en Amazon Redshift.

Cuando trabaje con la característica de consulta entre bases de datos de Amazon Redshift, tenga en cuenta lo siguiente:

- No puede consultar vistas creadas en otras bases de datos que hacen referencia a objetos de otra base de datos.
- No puede escribir en esquemas externos a objetos de base de datos en ninguna otra base de datos no conectada.
- Solo puede crear vistas materializadas y de enlace de tiempo de ejecución en objetos de otras bases de datos del clúster. No puede crear vistas regulares en objetos de otras bases de datos del clúster.
- Amazon Redshift no admite tablas con privilegios de nivel de columna para consultas entre bases de datos.
- No se admite la ejecución de consultas en varias bases de datos en tablas con claves de clasificación intercaladas.

# Ejemplos de consultas entre bases de datos

Este tema contiene ejemplos sobre cómo usar las consultas entre bases de datos. Las consultas entre bases de datos son consultas que funcionan en varias bases de datos dentro de un único clúster de Amazon Redshift.

Utilice los siguientes ejemplos para aprender a configurar una consulta entre bases de datos que haga referencia a una base de datos de Amazon Redshift.

Para comenzar, cree las bases de datos `db1` y `db2` y los usuarios `user1` y `user2` en el clúster de Amazon Redshift. Para obtener más información, consulte CREATE DATABASE y CREAR USUARIO.

```
--As user1 on db1
CREATE DATABASE db1;

CREATE DATABASE db2;

CREATE USER user1 PASSWORD 'Redshift01';

CREATE USER user2 PASSWORD 'Redshift01';
```

Como `user1` en `db1`, cree una tabla, conceda privilegios de acceso a `user2` e inserte valores en `table1`. Para obtener más información, consulte GRANT y INSERT.

```
--As user1 on db1
CREATE TABLE table1 (c1 int, c2 int, c3 int);

GRANT SELECT ON table1 TO user2;

INSERT INTO table1 VALUES (1,2,3),(4,5,6),(7,8,9);
```

Como `user2` en `db2`, ejecute una consulta entre bases de datos en `db2` mediante la notación de tres partes.

```
--As user2 on db2
SELECT * from db1.public.table1 ORDER BY c1;

c1 | c2  | c3
---+-----+----
```

```
1  |  2  | 3
4  |  5  | 6
7  |  8  | 9
(3 rows)
```

Ahora conceda privilegios de escritura a `user2` e inserte valores en `table1` en db1 como `user2`.

```
--As user1 on db1
GRANT INSERT ON table1 TO user2;
```

Como `user2` en db2, ejecute una consulta entre bases de datos en db2 mediante la notación de tres partes para insertar datos en `table1` en db1.

```
--As user2 on db2
INSERT INTO db1.public.table1 VALUES (10,11,12);
SELECT * from db1.public.table1 ORDER BY c1;

c1  | c2   | c3
----+------+----
1   |  2   | 3
4   |  5   | 6
7   |  8   | 9
10  |  11  | 12
(4 rows)
```

Como `user2` en db2, cree un esquema externo y ejecute una consulta entre bases de datos en db2 mediante la notación de esquema externo.

```
--As user2 on db2
CREATE EXTERNAL SCHEMA db1_public_sch
FROM REDSHIFT DATABASE 'db1' SCHEMA 'public';

SELECT * FROM db1_public_sch.table1 ORDER BY c1;

c1  | c2 | c3
----+----+----
1   | 2  | 3
4   | 5  | 6
7   | 8  | 9
10  | 11 | 12
(4 rows)
```

Para crear vistas diferentes y conceder permisos a esas vistas, como `user1` en `db1`, haga lo siguiente.

```
--As user1 on db1
CREATE VIEW regular_view AS SELECT c1 FROM table1;

GRANT SELECT ON regular_view TO user2;


CREATE MATERIALIZED VIEW mat_view AS SELECT c2 FROM table1;

GRANT SELECT ON mat_view TO user2;


CREATE VIEW late_bind_view AS SELECT c3 FROM public.table1 WITH NO SCHEMA BINDING;

GRANT SELECT ON late_bind_view TO user2;
```

Como `user2` en `db2`, ejecute la siguiente consulta entre bases de datos mediante la notación de tres partes para visualizar la vista en particular.

```
--As user2 on db2
SELECT * FROM db1.public.regular_view;

c1
----
1
4
7
10
(4 rows)

SELECT * FROM db1.public.mat_view;

c2
----
2
5
8
11
(4 rows)

SELECT * FROM db1.public.late_bind_view;
```

```
c3
----
3
6
9
12
(4 rows)
```

Como `user2` en `db2`, ejecute la siguiente consulta entre bases de datos mediante la notación de esquema externo para consultar la vista de enlace de tiempo de ejecución.

```
--As user2 on db2
SELECT * FROM db1_public_sch.late_bind_view;

c3
----
3
6
9
12
(4 rows)
```

Como `user2` en `db2`, ejecute el siguiente comando mediante tablas conectadas en una sola consulta.

```
--As user2 on db2
CREATE TABLE table1 (a int, b int, c int);

INSERT INTO table1 VALUES (1,2,3), (4,5,6), (7,8,9);

SELECT a AS col_1, (db1.public.table1.c2 + b) AS sum_col2, (db1.public.table1.c3 + c)
 AS sum_col3 FROM db1.public.table1, table1 WHERE db1.public.table1.c1 = a;
col_1 | sum_col2 | sum_col3
------+----------+----------
1     | 4        | 6
4     | 10       | 12
7     | 16       | 18
(3 rows)
```

En el siguiente ejemplo, se muestran todas las bases de datos del clúster.

```
select database_name, database_owner, database_type
from svv_redshift_databases
where database_name in ('db1', 'db2');

 database_name | database_owner | database_type
---------------+----------------+---------------
 db1           |            100 | local
 db2           |            100 | local
(2 rows)
```

En el siguiente ejemplo, se muestran todos los esquemas de Amazon Redshift de todas las bases de datos del clúster.

```
select database_name, schema_name, schema_owner, schema_type
from svv_redshift_schemas
where database_name in ('db1', 'db2');

 database_name |    schema_name     | schema_owner | schema_type
---------------+--------------------+--------------+-------------
 db1           | pg_catalog         |            1 | local
 db1           | public             |            1 | local
 db1           | information_schema |            1 | local
 db2           | pg_catalog         |            1 | local
 db2           | public             |            1 | local
 db2           | information_schema |            1 | local
(6 rows)
```

En el siguiente ejemplo, se muestran todas las tablas o las vistas de Amazon Redshift de todas las bases de datos del clúster.

```
select database_name, schema_name, table_name, table_type
from svv_redshift_tables
where database_name in ('db1', 'db2') and schema_name in ('public');

 database_name | schema_name |     table_name      | table_type
---------------+-------------+---------------------+------------
 db1           | public      | late_bind_view      | VIEW
 db1           | public      | mat_view            | VIEW
 db1           | public      | mv_tbl__mat_view__0 | TABLE
 db1           | public      | regular_view        | VIEW
 db1           | public      | table1              | TABLE
 db2           | public      | table2              | TABLE
```

```
(6 rows)
```

En el siguiente ejemplo, se muestran todos los esquemas externos y de Amazon Redshift de todas las bases de datos del clúster.

```
select database_name, schema_name, schema_owner, schema_type
from svv_all_schemas where database_name in ('db1', 'db2') ;

 database_name |    schema_name     | schema_owner | schema_type
---------------+--------------------+--------------+-------------
 db1           | pg_catalog         |            1 | local
 db1           | public             |            1 | local
 db1           | information_schema |            1 | local
 db2           | pg_catalog         |            1 | local
 db2           | public             |            1 | local
 db2           | information_schema |            1 | local
 db2           | db1_public_sch     |            1 | external
(7 rows)
```

En el siguiente ejemplo, se muestran todas las tablas externas y de Amazon Redshift de todas las bases de datos del clúster.

```
select database_name, schema_name, table_name, table_type
from svv_all_tables
where database_name in ('db1', 'db2') and schema_name in ('public');

 database_name | schema_name |     table_name      | table_type
---------------+-------------+---------------------+------------
 db1           | public      | regular_view        | VIEW
 db1           | public      | mv_tbl__mat_view__0 | TABLE
 db1           | public      | mat_view            | VIEW
 db1           | public      | late_bind_view      | VIEW
 db1           | public      | table1              | TABLE
 db2           | public      | table2              | TABLE
(6 rows)
```

# Uso de consultas entre bases de datos con el editor de consultas

Este tema explica cómo usar consultas entre bases de datos con el editor de consultas. Las consultas entre bases de datos son consultas que funcionan en varias bases de datos dentro de un único clúster de Amazon Redshift.

Puede utilizar las consultas entre bases de datos para acceder a los datos de cualquiera de las bases de datos del clúster de Amazon Redshift sin tener que conectarse a esa base de datos específica. Cuando ejecute consultas entre bases de datos en cualquier otra base de datos no conectada, solo tendrá acceso de lectura y escritura a esos objetos de la base de datos.

Puede consultar otros objetos de base de datos y escribir en ellos mediante nombres de objetos completos expresados con la notación de tres partes. La ruta completa a cualquier objeto de base de datos consta de tres componentes: el nombre de base de datos, el esquema y el nombre del objeto. Un ejemplo es *`database_name.schema_name.object_name`*.

Para usar consultas entre bases de datos con el editor de consultas v2

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. Cree un clúster para utilizar consultas entre bases de datos en el editor de consultas v2 de Amazon Redshift. Para obtener más información, consulte Creación de un clúster en la Guía de administración de Amazon Redshift.
3. Habilite el acceso al editor de consultas con los permisos necesarios. Para obtener más información, consulte Consulta de una base de datos mediante el editor de consultas v2 en la Guía de administración de Amazon Redshift.
4. En el menú de navegación, elija Editor de consultas v2 y conéctese a una base de datos en el clúster.

   Cuando se conecte al editor de consultas v2 por primera vez, Amazon Redshift mostrará los recursos de la base de datos conectada de forma predeterminada.
5. Elija las otras bases de datos a las que tiene acceso para ver los objetos de base de datos de estas otras bases de datos. Asegúrese de tener los permisos necesarios para ver los objetos. Una vez que eligió una base de datos, Amazon Redshift muestra la lista de esquemas de la base de datos.

   Seleccione un esquema para ver la lista de objetos de base de datos dentro de ese esquema.

> **Note**
>
> Amazon Redshift no admite de forma directa los objetos de catálogo de consultas que forman parte de AWS Glue o bases de datos federadas. Para consultar estos objetos, primero cree esquemas externos que hagan referencia a estos orígenes de datos externos en cada base de datos.

Las consultas entre bases de datos de Amazon Redshift con notación de tres partes no admiten tablas de metadatos en los esquemas `information_schema` y `pg_catalog`, ya que estas vistas de metadatos son específicas de una base de datos.

6. (Opcional) Filtre la lista de tablas o de vistas para el esquema seleccionado.

# Compatibilidad de Apache Iceberg con Amazon Redshift

Puede registrar clústeres aprovisionados o espacios de nombres sin servidor completos de Amazon Redshift en el AWS Glue Data Catalog para crear catálogos que compartan de forma segura los datos en directo en todas las cuentas de AWS. Puede acceder a estos catálogos desde cualquier motor de consultas de SQL que sea compatible con la API de REST de Apache Iceberg. AWS Lake Formation administra los permisos de los catálogos, lo que le permite administrar una sola copia de los datos con un único conjunto de permisos y, al mismo tiempo, aprovechar las características de Amazon Redshift, como las vistas materializadas y las integraciones sin ETL.

Todos los catálogos creados a partir de clústeres aprovisionados y espacios de nombres sin servidor registrados de Amazon Redshift en el AWS Glue Data Catalog se montan automáticamente como bases de datos externas en todos los clústeres aprovisionados y grupos de trabajo sin servidor en la misma Región de AWS de la misma cuenta. Los catálogos creados en el AWS Glue Data Catalog para almacenar datos en Redshift Managed Storage (RMS) se montan de forma similar a las bases de datos externas. Una vez montadas, puede conectarse directamente a estas bases de datos y consultar los objetos compartidos utilizando la notación de tres partes `database@namespace-catalog.schema.table`.

## Regiones en las que está disponible la compatibilidad con Apache Iceberg

La compatibilidad de Apache Iceberg con Amazon Redshift está disponible en las siguientes Regiones de AWS:

- Este de EE. UU. (Norte de Virginia)
- Este de EE. UU. (Ohio)
- Oeste de EE. UU. (Norte de California)
- Asia-Pacífico (Hong Kong)
- Asia-Pacífico (Seúl)
- Asia-Pacífico (Singapur)
- Asia-Pacífico (Sídney)
- Asia-Pacífico (Tokio)
- Canadá (centro)

- Europa (Fráncfort)
- Europa (Irlanda)
- Europe (Londres)
- Europe (Stockholm)
- América del Sur (São Paulo)

# Consideraciones y limitaciones al usar catálogos de Amazon Redshift en el AWS Glue Data Catalog

Cuando utilice catálogos de Amazon Redshift en AWS Glue Data Catalog, tenga en cuenta lo siguiente:

- Los almacenamientos de datos registrados en AWS Glue Data Catalog siguen la sintaxis de tres partes para acceder a las tablas (`database@namespace-catalog.schema.table`). Por ejemplo, si ha registrado un espacio de nombres de Amazon Redshift denominado a, que ha rellenado una base de datos denominada b, que contenía un esquema denominado c, en el que había una tabla denominada d, seleccionaría una de las opciones d mediante la siguiente instrucción:

  ```
  SELECT * FROM b@a.c.d;
  ```

  Tenga en cuenta que la longitud combinada de la parte `database@namespace-catalog` de la sintaxis debe ser de 127 caracteres o menos.
- Al registrar un clúster o un espacio de nombres en el AWS Glue Data Catalog, Amazon Redshift registra todas las bases de datos y las relaciones de ese clúster o espacio de nombres.
- Puede registrar varios clústeres y espacios de nombres de Redshift en el AWS Glue Data Catalog.
- Al registrar un clúster o espacio de nombres, solo se registran los esquemas y relaciones internos de ese clúster o espacio de nombres. No se registra lo siguiente:
  - Esquemas externos
  - Tablas externas. Tenga en cuenta que se registrarán las vistas de enlace tardío creadas a partir de tablas externas.
  - Funciones creadas por el usuario
  - Procedimientos
  - Tablas con políticas de seguridad de fila o de enmascaramiento de datos dinámico asociadas.

  - Objetos de base de datos con nombres en mayúsculas o en mayúsculas y minúsculas. Las tablas que contienen columnas en mayúsculas o en mayúsculas y minúsculas no se registran. Esto se aplica incluso cuando enable_case_sensitive_identifier está deshabilitado.
- Los permisos de base de datos de Amazon Redshift, como los roles que concede el control de acceso basado en roles, no se transfieren a los catálogos del AWS Glue Data Catalog. Utilice AWS Lake Formation para configurar permisos para el AWS Glue Data Catalog. Para obtener más información sobre el uso de Lake Formation para configurar los permisos, consulte Configurar permisos en recursos compartidos de datos de Amazon Redshift en la Guía para desarrolladores de AWS Lake Formation.
- Al crear un catálogo a partir de un clúster registrado o un espacio de nombres sin servidor, el AWS Glue Data Catalog crea un grupo de trabajo administrado de Amazon Redshift con los recursos de computación de Amazon Redshift para gestionar los requisitos de computación para consultar ese catálogo. Puede ver el grupo de trabajo administrado en la consola de Amazon Redshift sin servidor y administrarlo en AWS Glue.
- Cuando registre un clúster en pausa, el AWS Glue Data Catalog no lo montará como catálogo hasta que se reanude.
- Al registrar un espacio de nombres sin servidor que no se esté utilizando de forma activa, el AWS Glue Data Catalog no lo montará como catálogo hasta que se vuelva a utilizar.
- Su cuenta debe tener una VPC predeterminada para crear un grupo de trabajo administrado.

# Requisitos de la política de IAM para acceder a los clústeres y espacios de nombres registrados en el Catálogo de datos

En este tema se describen los permisos de IAM necesarios para registrar los clústeres aprovisionados y los espacios de nombres sin servidor en el Catálogo de datos y acceder a ellos con Amazon Redshift.

Después de registrar un clúster aprovisionado o un espacio de nombres sin servidor en el AWS Glue Data Catalog, necesitará los siguientes permisos para detectar la creación del catálogo y los cambios realizados en él posteriormente.

- `glue:GetCatalog`
- `glue:GetCatalogs`
- `lakeformation:GetDataAccess`

Estos permisos se incluyen en el rol vinculado al servicio `AmazonRedshiftServiceLinkedRolePolicy`. Para obtener más información sobre este rol, consulte Uso de roles vinculados a servicios para Amazon Redshift en la Guía de administración de Amazon Redshift.

# Registro de clústeres y espacios de nombres de Amazon Redshift en el AWS Glue Data Catalog

Puede añadir clústeres aprovisionados y espacios de nombres sin servidor de Amazon Redshift al AWS Glue Data Catalog para acceder a ellos mediante la API de REST de Apache Iceberg. Para ello, registre el almacenamiento de datos de Amazon Redshift en el AWS Glue Data Catalog utilizando la consola de Amazon Redshift o la AWS CLI y, a continuación, cree un catálogo federado de Amazon Redshift para el almacenamiento que utiliza AWS Lake Formation.

Los almacenamientos de datos de Amazon Redshift registrados en el catálogo de datos actúan como recursos compartidos de datos del productor. Los cambios que realice en los clústeres o los espacios de nombres sin servidor desde sus catálogos se reflejan en el clúster o el espacio de nombres de Redshift y viceversa.

Registering using the Amazon Redshift console

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. Vaya hasta el espacio de nombres sin servidor o el clúster aprovisionado que desee registrar y selecciónelo.
3. En la página de detalles del espacio de nombres o del clúster, seleccione Registrar en AWS Glue Data Catalog en el menú desplegable Acciones.
4. Introduzca el ID de cuenta de destino en la que desea registrar el espacio de nombres o el clúster y seleccione Registrar.
5. Si va a registrarse desde su cuenta en la misma cuenta de AWS Glue, accederá directamente a la consola de AWS Lake Formation para finalizar el proceso. Si va a registrarse en una cuenta diferente, aparecerá un enlace que le llevará a Lake Formation.

Registering using the AWS CLI

Para registrar un clúster o un espacio de nombres en el AWS Glue Data Catalog mediante la AWS CLI, utilice el comando `register-namespace` con las siguientes opciones:

- `namespace-identifier`: un objeto con el identificador único del clúster o espacio de nombres que va a registrar. Este objeto es diferente en función de si va a registrar un clúster aprovisionado o un espacio de nombres sin servidor. Considere lo siguiente:
  - En el caso de los clústeres aprovisionados, se proporciona un objeto `ProvisionedIdentifier` que contiene un objeto `ClusterIdentifier` con el identificador único del clúster que se va a registrar.
  - En el caso de los espacios de nombres sin servidor, se proporciona un objeto `ServerlessIdentifier` que contiene un objeto `NamespaceIdentifier` con el identificador único del espacio de nombres que se va a registrar, así como un objeto `WorkgroupIdentifier` con el identificador único del grupo de trabajo asociado a ese espacio de nombres.
- `consumer-identifiers`: una matriz con un solo elemento que contiene el identificador único de la cuenta en la que va a registrar el clúster o el espacio de nombres.

En el siguiente ejemplo, se registra el espacio de nombres sin servidor `mySampleNamespace` en el ID de la cuenta. `012345678910`.

```
aws redshift register-namespace /
--namespace-identifier {ServerlessIdentifier: {NamespaceIdentifer:
 mySampleNamespace, WorkgroupIdentifier: mySampleWorkgroup}} /
--consumer-identifiers [012345678910]
```

# Anulación del registro de clústeres y espacios de nombres de Amazon Redshift del AWS Glue Data Catalog

Puede anular el registro de un clúster aprovisionado o un espacio de nombres sin servidor desde el AWS Glue Data Catalog mediante la consola de Amazon Redshift o la AWS CLI. Los almacenamientos de datos en los que se anula el registro del AWS Glue Data Catalog solo se eliminan de AWS Glue Data Catalog. Los almacenamientos siguen formando parte de la cuenta de Amazon Redshift. Además, el catálogo permanece en AWS Glue y debe eliminarse manualmente. Al eliminar el catálogo de AWS Glue, también se elimina el grupo de trabajo administrado asociado a él.

Deregistering using the Amazon Redshift console

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.

2. Vaya al espacio de nombres sin servidor o el clúster aprovisionado del que desee anular el registro y selecciónelo.
3. En la página de detalles del espacio de nombres o del clúster, seleccione Anular el registro de AWS Glue Data Catalog en el menú desplegable Acciones. Esta opción solo aparece si ha elegido un almacenamiento de datos que ya esté registrado en el AWS Glue Data Catalog.
4. Introduzca el ID de cuenta propietaria del catálogo que contiene el clúster o espacio de nombres y elija Anular registro.

Deregistering using the AWS CLI

Para anular el registro de un clúster o un espacio de nombres del AWS Glue Data Catalog mediante la AWS CLI, utilice el comando `deregister-namespace` con las siguientes opciones:

- `namespace-identifier`: un objeto con el identificador único del clúster o espacio de nombres cuyo registro está anulando. Este objeto es diferente en función de si va a anular el registro de un clúster aprovisionado o un espacio de nombres sin servidor. Considere lo siguiente:
  - En el caso de los clústeres aprovisionados, se proporciona un objeto `ProvisionedIdentifier` que contiene un objeto `ClusterIdentifier` con el identificador único del clúster cuyo registro se va a anular.
  - En el caso de los espacios de nombres sin servidor, se proporciona un objeto `ServerlessIdentifier` que contiene un objeto `NamespaceIdentifier` con el identificador único del espacio de nombres cuyo registro se va a anular, así como un objeto `WorkgroupIdentifier` con el identificador único del grupo de trabajo asociado a ese espacio de nombres.
- `consumer-identifiers`: una matriz con un solo elemento que contiene el identificador único de la cuenta en la que se va a anular el registro del clúster o el espacio de nombres.

En el siguiente ejemplo, se anula el registro del espacio de nombres sin servidor `mySampleNamespace` del ID de cuenta `012345678910`.

```
aws redshift deregister-namespace /
--namespace-identifier {ServerlessIdentifier: {NamespaceIdentifer:
 mySampleNamespace, WorkgroupIdentifier: mySampleWorkgroup}} /
--consumer-identifiers [012345678910]
```

# Grupos de trabajo administrados

Al registrar un clúster aprovisionado o un espacio de nombres sin servidor en el AWS Glue Data Catalog y crear un catálogo a partir de él, AWS Glue crea un grupo de trabajo administrado para proporcionar recursos de computación a cualquier motor de consultas de SQL que acceda a ese catálogo. Por ejemplo, si un usuario de Amazon Athena consulta una tabla de un catálogo, los recursos de computación de Amazon Redshift del grupo de trabajo administrado cubrirían sus requisitos de computación. Los grupos de trabajo administrados utilizan los recursos de computación de Amazon Redshift, pero se administran en AWS Glue. Su cuenta debe tener una VPC predeterminada para crear un grupo de trabajo administrado.

Los grupos de trabajo administrados se incluyen en la cuota de grupos de trabajo de Amazon Redshift sin servidor, que de forma predeterminada es 25. Para obtener más información, consulte Cuotas y límites de Amazon Redshift en la Guía de administración de Amazon Redshift.

Puede ver información general, métricas de rendimiento e información de consultas de sus grupos de trabajo administrados en la consola de Amazon Redshift sin servidor. Para obtener más información, consulte Visualización de las propiedades de un grupo de trabajo en la Guía de administración de Amazon Redshift.

# Consulta de catálogos registrados en el AWS Glue Data Catalog

Tras registrar un almacenamiento de datos de Amazon Redshift en el AWS Glue Data Catalog y configurar los permisos para el catálogo resultante en AWS Lake Formation, el catálogo se monta automáticamente en todas las instancias de Amazon Redshift con acceso al almacenamiento de datos de origen en la misma cuenta y Región de AWS. A continuación, puede consultar ese catálogo como lo haría con un clúster o grupo de trabajo local. También puede consultar los catálogos registrados en el AWS Glue Data Catalog mediante cualquier motor de SQL que sea compatible con la API abierta de REST de Apache Iceberg. Para obtener más información sobre la consulta de catálogos en AWS Glue Data Catalog mediante la API de REST de Apache Iceberg, consulte Cómo acceder al Catálogo de datos en la Guía para desarrolladores de AWS Glue. Para obtener información sobre la API de REST de Apache Iceberg, consulte la especificación de la API abierta de REST de Apache Iceberg.

Para consultar un catálogo, primero debe configurar los permisos de ese catálogo mediante AWS Lake Formation. Para obtener más información sobre la configuración de permisos para catálogos en AWS Lake Formation, consulte Configurar permisos en recursos compartidos de datos de Amazon Redshift en la Guía para desarrolladores de AWS Lake Formation. También necesita un rol de

IAM con la política administrada `AmazonRedshiftServiceLinkedRolePolicy` asociada. Para obtener más información sobre los roles vinculados al servicio, consulte Uso de roles vinculados a servicios para Amazon Redshift en la Guía de administración de Amazon Redshift.

Tenga en cuenta que las consultas a los catálogos deben utilizar la siguiente sintaxis de tres partes para acceder a las tablas:

```
database@namespace.schema.table
```

Para obtener información general sobre las consultas de almacenamientos de datos de Amazon Redshift, visite Consulta de una base de datos en la Guía de administración de Amazon Redshift.

Querying using the query editor v2

Tras configurar los permisos para que una cuenta acceda a un grupo de trabajo administrado, dicho grupo de trabajo administrado aparece en el panel de vista de árbol situado en la sección de bases de datos externas de la base de datos sin servidor. Puede consultar el grupo de trabajo administrado de la misma manera que consultaría un clúster interno aprovisionado o un grupo de trabajo sin servidor de Amazon Redshift, utilizando el formato de sintaxis de tres partes `database@namespace/cluster.schema.table`. Consulte la siguiente instrucción de ejemplo:

```
SELECT price FROM sales_db@mynamespace.sales_schema.inventory_table
```

Querying using the Data API

Puede consultar los grupos de trabajo administrados mediante la API de datos de Amazon Redshift del mismo modo que consultaría un clúster aprovisionado o un grupo de trabajo sin servidor internos de Amazon Redshift, pasando el nombre de recurso de Amazon (ARN) del catálogo al atributo `database`correspondiente. Considere el siguiente ejemplo, en el que se crea una tabla en un catálogo.

```
aws redshift-data execute-statement –sql 'CREATE TABLE IF NOT EXISTS "dev@test-
rms-catalog".public.t1 (c1 INT, c2 VARCHAR(10));' –database arn:aws:glue:us-
east-1:550022730026:catalog/test-rms-catalog
```

# Uso compartido de datos en Amazon Redshift

Con Amazon Redshift, puede compartir datos de forma segura entre clústeres de Amazon Redshift o con otros servicios de AWS. El uso compartido de datos le permite compartir datos en tiempo real, sin tener que crear una copia ni moverlos. Los administradores de bases de datos y los ingenieros de datos pueden utilizar el intercambio de datos para proporcionar un acceso seguro y de solo lectura a los datos con fines de análisis y, al mismo tiempo, mantener el control sobre los datos. Los analistas de datos, los profesionales de inteligencia empresarial y los científicos de datos pueden aprovechar los datos compartidos para obtener información sin duplicarlos ni moverlos. Los casos de uso comunes incluyen compartir datos con socios, permitir el análisis multifuncional y facilitar la democratización de los datos dentro de una organización. En las siguientes secciones se describen los detalles de la configuración y la administración del uso compartido de datos en Amazon Redshift.

Con el uso compartido de datos de Amazon Redshift, puede compartir de forma segura el acceso a datos en directo en clústeres de Amazon Redshift, grupos de trabajo, Cuentas de AWS y Regiones de AWS sin mover ni copiar los datos manualmente. Desde que los datos son en directo, todos los usuarios pueden ver la información más actualizada y coherente en Amazon Redshift en cuanto se actualiza.

Puede compartir datos entre clústeres aprovisionados, grupos de trabajo sin servidor, zonas de disponibilidad, Cuentas de AWS y Regiones de AWS. Puede compartir entre tipos de clústeres, así como entre clústeres aprovisionados y sin servidor.

Puede compartir objetos de base de datos para lecturas y escrituras entre distintos clústeres de Amazon Redshift o grupos de trabajo de Amazon Redshift sin servidor dentro de la misma Cuenta de AWS o de una Cuenta de AWS a otra. También puede leer y escribir datos entre regiones. Puede conceder permisos como SELECT, INSERT y UPDATE para distintas tablas y USAGE y CREATE para distintos esquemas. Los datos están activos y disponibles para todos los almacenes en cuanto se confirma una transacción de escritura.

## Casos de uso compartido de datos en Amazon Redshift

El uso compartido de datos de Amazon Redshift es especialmente útil para estos casos de uso:

- Compatibilidad con diferentes tipos de cargas de trabajo críticas para la empresa: utilice un clúster central de extracción, transformación y carga (ETL) que comparta los datos con varios clústeres de análisis o inteligencia empresarial (BI). Este enfoque proporciona aislamiento de la carga de

trabajo de lectura y reintegro para las cargas de trabajo individuales. Puede ajustar el tamaño y la escala de la informática de la carga de trabajo individual de acuerdo con los requisitos de precio y rendimiento específicos de la carga de trabajo.

- Habilitación de la colaboración entre grupos: habilite la colaboración continua entre los equipos y los grupos empresariales para lograr analítica, ciencia de datos y análisis de impacto entre productos más amplios.
- Entrega de datos como servicio: comparta datos como un servicio con toda la organización.
- Uso compartido de datos entre entornos: comparta datos entre entornos de desarrollo, prueba y producción. Puede mejorar la agilidad del equipo compartiendo datos con diferentes niveles de detalle.
- Acceso con licencia a los datos de Amazon Redshift: enumere los conjuntos de datos de Amazon Redshift en el catálogo de AWS Data Exchange que los clientes pueden encontrar y consultar, además de suscribirse a ellos, en cuestión de minutos.

## Casos de uso compartido de datos para accesos de escritura

El uso compartido de datos para escrituras tiene varios casos de uso importantes:

- Actualizar los datos de origen empresarial en el productor: puede compartir los datos como un servicio en toda su organización, pero los consumidores también pueden realizar acciones en los datos de origen. Por ejemplo, pueden comunicar valores actualizados o confirmar la recepción de los datos. Estos son solo un par de posibles casos de uso empresariales.
- Insertar registros adicionales en el productor: los consumidores pueden agregar registros a los datos de origen originales. Si es necesario, se pueden marcar como procedentes del consumidor.

Para obtener información específica sobre cómo realizar operaciones de escritura en un recurso compartido de datos, consulte Uso compartido del acceso de escritura a los datos.

# Uso compartido de datos en niveles diferentes en Amazon Redshift

Con Amazon Redshift, puede compartir datos en niveles diferentes. Estos niveles incluyen las bases de datos, los esquemas, las tablas, las vistas (incluidas las vistas normales, de enlace de tiempo de ejecución y materializadas) y las funciones definidas por el usuario (UDF) de SQL. Puede crear varios datashares para determinada base de datos. Un datashare puede contener objetos de varios esquemas de la base de datos en la cual se crea el uso compartido.

Contar con esta flexibilidad para compartir datos le permite obtener control de acceso pormenorizado. Puede personalizar este control para diferentes usuarios y empresas que necesiten acceder a los datos de Amazon Redshift.

# Administración de la coherencia del uso compartido de datos en Amazon Redshift

Amazon Redshift proporciona consistencia transaccional en todos los clústeres productores y consumidores, y comparte vistas actualizadas y consistentes de los datos con todos los consumidores.

Los datos se pueden actualizar de forma continua en el clúster productor. Todas las consultas de un clúster consumidor dentro de una transacción leen el mismo estado de los datos compartidos. Amazon Redshift no tiene en cuenta los datos modificados por otra transacción del clúster productor que se haya confirmado después de que se inicia la transacción en el clúster consumidor. Después de confirmar el cambio en los datos en el clúster productor, las transacciones nuevas que se produzcan en el clúster consumidor podrán consultar de inmediato los datos actualizados.

La consistencia sólida elimina el riesgo de que se generen informes empresariales poco confiables que puedan contener resultados no válidos mientras se comparten los datos. Este factor es especialmente importante para el análisis financiero o para los casos en que los resultados puedan utilizarse para preparar conjuntos de datos que se utilicen para formar modelos de machine learning.

# Consideraciones sobre el uso compartido de datos en Amazon Redshift

Con el uso compartido de datos de Amazon Redshift, puede compartir de forma segura el acceso a datos en directo en clústeres de Amazon Redshift, grupos de trabajo, Cuentas de AWS y Regiones de AWS sin mover ni copiar los datos manualmente. Anteriormente, los objetos de los recursos compartidos de datos eran de solo lectura en todas las circunstancias. Escribir en un objeto de un recurso compartido de datos es una característica nueva. Los objetos de los recursos compartidos de datos solo están habilitados para escritura cuando un productor concede específicamente privilegios de escritura, como INSERT o CREATE, a los objetos del recurso compartido de datos. Además, para el uso compartido entre cuentas, el productor debe autorizar las escrituras en el recurso compartido de datos y el consumidor debe asociar clústeres y grupos de trabajo específicos para las escrituras.

En esta sección, se indican los aspectos que se deben tener en cuenta al trabajar con recursos compartidos de datos de Amazon Redshift.

## Consideraciones generales sobre el uso compartido de datos en Amazon Redshift

Estos son los aspectos generales que se deben tener en cuenta al trabajar con recursos compartidos de datos en Amazon Redshift.

- Base de datos predeterminada: al leer datos de un recurso compartido de datos, permanece conectado a la base de datos del clúster local. Para obtener más información sobre cómo establecer a y leer de una base de datos creada a partir de un recurso compartido de datos, consulte Consulta de objetos de recursos compartidos de datos y Vistas materializadas de tablas de lagos de datos externos en Amazon Redshift Spectrum.
- Conexiones: debe estar conectado directamente a una base de datos del recurso compartido de datos o ejecutar el comando USE para escribir en recursos compartidos de datos. También puede utilizar la notación de tres partes. El comando USE no se admite en tablas externas.
- Rendimiento: el rendimiento de las consultas sobre datos compartidos depende de la capacidad de computación de los clústeres consumidores.
- Cargos por transferencia de datos: el uso compartido de datos entre regiones conlleva cargos adicionales por transferencia de datos entre regiones.
  - Estos cargos por transferencia de datos no se aplican dentro de la misma región, solo entre regiones. Para obtener más información, consulte Administración del control de costos para el uso compartido de datos entre regiones.
  - Al consumidor se le cobran todas las tarifas informáticas y de transferencia de datos entre regiones necesarias para consultar los datos del productor. Al productor se le cobra por el almacenamiento subyacente de los datos en el clúster aprovisionado o en el espacio de nombres sin servidor.
- Uso compartido de datos dentro de los clústeres y entre ellos: solo necesita los recursos compartidos de datos al compartir datos entre distintos clústeres aprovisionados o grupos de trabajo sin servidor de Amazon Redshift. Dentro del mismo clúster, puede consultar otra base de datos mediante una notación simple de tres partes `database.schema.table`, siempre que disponga de los permisos necesarios en los objetos de la otra base de datos.
- Descubrimiento de metadatos: si es un consumidor conectado directamente a una base de datos de recurso compartido de datos mediante los controladores Redshift JDBC, ODBC o Python, puede ver los datos del catálogo de las siguientes maneras:

  - Los comandos SHOW de SQL
  - Consulta de las tablas y vistas information_schema
  - Consulta de las vistas de metadatos SVV
- Visibilidad de permisos: los consumidores pueden ver los permisos concedidos a los recursos compartidos de datos a través del comando de SQL SHOW GRANTS.
- Administración del cifrado de clústeres para el uso compartido de datos: para compartir datos entre una Cuenta de AWS, tanto el clúster productor como el consumidor deben estar cifrados.
  - Si los clústeres del productor y del consumidor y los espacios de nombres sin servidor están en la misma cuenta, deben tener el mismo tipo de cifrado (ambos sin cifrar o ambos cifrados). En todos los demás casos, como los recursos compartidos de datos administrados por Lake Formation, el consumidor y el productor deben estar cifrados. Esto se hace por motivos de seguridad. Sin embargo, no es necesario que compartan la misma clave de cifrado.
  - Para proteger los datos en tránsito, todos ellos se cifran en tránsito a través del esquema de cifrado del clúster productor. El clúster consumidor adopta este esquema de cifrado cuando se cargan los datos. Luego, el clúster consumidor funciona como un clúster cifrado normal. Las comunicaciones entre el productor y el consumidor también se cifran con un esquema de clave compartida. Para obtener más información sobre el cifrado en tránsito, consulte Cifrado en tránsito.

## Consideraciones sobre lecturas y escrituras mediante uso compartido de datos en Amazon Redshift

> **Note**
>
> Las escrituras en varios almacenamientos de Amazon Redshift mediante el uso compartido de datos solo se admiten en el parche 186 de Amazon Redshift para los clústeres aprovisionados en la versión 1.0.78881 actual o posteriores y para los grupos de trabajo de Amazon Redshift sin servidor en la versión 1.0.78890 o posteriores.

Estos son los aspectos que se deben tener en cuenta al trabajar con lecturas y escrituras mediante el uso compartido de datos en Amazon Redshift:

- Solo puede compartir UDF de SQL a través de recursos compartidos de datos. No se admiten los UDF de Python y Lambda.

- Si la base de datos productora tiene una intercalación específica, utilice la misma configuración de intercalación para la base de datos consumidora.
- Amazon Redshift no admite funciones definidas por el usuario de SQL anidadas en clústeres productores.
- Amazon Redshift no admite el uso compartido de tablas con claves de ordenación intercaladas ni vistas que referencien tablas con claves de ordenación intercaladas.
- Amazon Redshift no admite el acceso a un objeto de recurso compartido de datos en el que se haya producido un DDL simultáneo entre la preparación y la ejecución del acceso.
- Amazon Redshift no admite el uso compartido de procedimientos almacenados a través de recursos compartidos de datos.
- Amazon Redshift no permite compartir tablas del sistema ni vistas del sistema de metadatos.
- Tipo de computación: debe utilizar grupos de trabajo sin servidor y clústeres ra3.large, ra3.xlplus, ra3.4xl o ra3.16xl para usar esta característica.
- Nivel de aislamiento: el nivel de aislamiento de la base de datos debe ser de instantáneas para permitir que otros grupos de trabajo sin servidor y clústeres aprovisionados escriban en ella.
- Consultas y transacciones de varias instrucciones: actualmente no se admiten las consultas de varias instrucciones fuera de un bloque de transacciones. Por tanto, si utiliza un editor de consultas como dbeaver y tiene varias consultas de escritura, debe incluir las consultas en una instrucción de transacción explícita BEGIN...END.

  Cuando se utilizan instrucciones de varios comandos fuera de las transacciones, si el primer comando es una escritura en una base de datos de productor, los siguientes comandos de escritura de la instrucción solo se permiten en esa base de datos de productor. Si el primer comando es una lectura, los comandos de escritura posteriores solo se permiten en la base de datos utilizada, si esta está configurada; de lo contrario, en la base de datos local. Tenga en cuenta que las escrituras en una transacción solo se admiten en una única base de datos.
- Dimensionamiento del consumidor: los clústeres del consumidor deben tener al menos 64 sectores para poder realizar escrituras mediante el uso compartido de datos.
- Vistas y vistas materializadas: no puede crear, actualizar ni alterar vistas ni vistas materializadas en una base de datos de recursos compartidos de datos.
- Seguridad: no puede asociar ni eliminar políticas de seguridad, como las de columna (CLS), fila (RLS) y enmascaramiento dinámico de datos (DDM), en los objetos de los recursos compartidos de datos.

- Facilidad de administración: los almacenes de consumidor no pueden añadir a otro recurso compartido de datos objetos de recursos compartidos de datos ni vistas que hagan referencia a objetos de recursos compartidos de datos. Los consumidores tampoco pueden modificar ni eliminar un recurso compartido de datos existente.
- Operaciones de truncado: las escrituras de recursos compartidos de datos admiten elementos truncados transaccionales para tablas remotas. Es diferente de los truncados que ejecuta localmente en un clúster, que son de confirmación automática. Para obtener más información sobre el comando SQL, consulte TRUNCATE.

## Consideraciones sobre el uso compartido de datos con tablas de lago de datos en Amazon Redshift

A continuación, se indican los aspectos que se deben tener en cuenta al trabajar con tablas de lago de datos en Amazon Redshift:

- Los recursos compartidos de tablas de lagos de datos no admiten AWS KMS keys administradas por el cliente para el cifrado de buckets de Amazon S3. Se puede utilizar Claves administradas por AWS para el cifrado. Para obtener más información, consulte Uso del cifrado del servidor con claves administradas por Amazon S3 (SSE-S3).
- Para compartir datos de tablas de lagos de datos de un catálogo cifrado del AWS Glue, debe delegar las operaciones de AWS KMS en un rol de IAM siguiendo las instrucciones de Cifrado del Catálogo de datos.

## Consideraciones sobre el uso compartido de datos con AWS Lake Formation en Amazon Redshift

A continuación, se indican las consideraciones y las limitaciones para compartir datos de Amazon Redshift con Lake Formation. Para obtener información sobre las consideraciones y limitaciones del uso compartido de datos, consulte Consideraciones sobre el uso compartido de datos en Amazon Redshift. Para obtener información sobre las limitaciones de Lake Formation, consulte Notas sobre cómo trabajar con los recursos compartidos de datos de Amazon Redshift en Lake Formation.

- Actualmente, no se permite compartir un recurso compartido de datos con Lake Formation entre varias regiones.

- Si se definen filtros a nivel de columna para un usuario en una relación compartida, la ejecución de una operación `SELECT *` devuelve solo las columnas a las que el usuario tiene acceso.
- No se admiten los filtros de nivel de celda de Lake Formation.
- Si creó y compartió una vista y sus tablas con Lake Formation, puede configurar filtros para administrar el acceso de las tablas. Amazon Redshift aplica las políticas definidas por Lake Formation cuando los usuarios del clúster consumidor acceden a los objetos compartidos. Cuando un usuario accede a una vista compartida con Lake Formation, Redshift aplica solo las políticas de Lake Formation definidas en la vista y no las tablas que contiene. Sin embargo, cuando los usuarios acceden directamente a la tabla, Redshift aplica las políticas de Lake Formation definidas en la tabla.
- No puede crear vistas materializadas sobre el consumidor en función de una tabla compartida si esta tiene configurados filtros de Lake Formation.
- El administrador de Lake Formation debe tener permisos de administrador del lago de datos y los permisos necesarios para aceptar un recurso compartido de datos.
- El clúster consumidor del productor debe ser un clúster RA3 con la última versión del clúster de Amazon Redshift o un grupo de trabajo sin servidor para compartir los recursos compartidos de datos a través de Lake Formation.
- Tanto el clúster productor como el consumidor deben estar cifrados.
- Las políticas de control de acceso en el nivel de fila y columna de Redshift implementadas en el clúster o grupo de trabajo del productor se ignoran cuando el recurso compartido de datos se comparte con Lake Formation. El administrador de Lake Formation debe configurar estas políticas en Lake Formation. El administrador del clúster o grupo de trabajo del productor puede desactivar RLS de una tabla mediante el comando ALTER TABLE.
- Compartir los recursos compartidos de datos a través de Lake Formation solo está disponible para los usuarios que tienen acceso a Redshift y Lake Formation.

## Consideraciones sobre el uso compartido de datos con AWS Data Exchange en Amazon Redshift

Cuando utilice AWS Data Exchange for Amazon Redshift, tenga en cuenta lo siguiente:

- Tanto los productores como los consumidores deben utilizar los tipos de instancias RA3 para utilizar los recursos compartidos de datos de Amazon Redshift. Los productores deben utilizar los tipos de instancias RA3 con la última versión del clúster de Amazon Redshift.

- Tanto el clúster productor como el consumidor deben estar cifrados.
- Debe estar registrado como proveedor de AWS Data Exchange para publicar productos en AWS Data Exchange, incluidos los productos que contienen recursos compartidos de datos de AWS Data Exchange. Para obtener más información, consulte Primeros pasos como proveedor.
- No es necesario que esté registrado como proveedor de AWS Data Exchange para buscar y consultar datos de Amazon Redshift, además de suscribirse a ellos, a través de AWS Data Exchange.
- Para controlar el acceso a los datos, cree recursos compartidos de datos de AWS Data Exchange con la configuración de acceso público activada. Para modificar un recurso compartido de datos de AWS Data Exchange para desactivar la configuración de acceso público, configure la variable de sesión para habilitar ALTER DATASHARE SET PUBLICACCESSIBLE FALSE. Para obtener más información, consulte Notas de uso de ALTER DATASHARE.
- Los productores no pueden agregar o eliminar de forma manual consumidores de los recursos compartidos de datos de AWS Data Exchange porque el acceso a estos recursos se otorga en función de si se tiene una suscripción activa a un producto de AWS Data Exchange que contiene el recurso compartido de datos de AWS Data Exchange.
- Los productores no pueden ver las consultas SQL que ejecutan los consumidores. Solo pueden ver los metadatos, como la cantidad de consultas o los objetos que consultan los consumidores, a través de tablas de Amazon Redshift a las que solo puede acceder el productor. Para obtener más información, consulte Monitoreo y auditoría en el uso compartido de datos en Amazon Redshift.
- Le recomendamos que sus recursos compartidos de datos sean accesibles públicamente. Si esto no sucede, los suscriptores de AWS Data Exchange con clústeres consumidores de acceso público no podrán utilizar su recurso compartido de datos.
- Le recomendamos que no elimine un recurso compartido de datos de AWS Data Exchange compartido con otras Cuentas de AWS mediante la instrucción DROP DATASHARE. Si lo hace, las Cuentas de AWS que tengan acceso al recurso compartido de datos perderán el acceso. Esta acción es irreversible. Realizar este tipo de alteración puede infringir los términos del producto de datos en AWS Data Exchange. Si desea borrar un recurso compartido de datos de AWS Data Exchange, consulte Notas de uso de DROP DATASHARE.
- Para compartir datos entre regiones, puede crear recursos compartidos de datos de AWS Data Exchange con objeto de compartir datos con licencia.
- Al consumir datos de una región diferente, el consumidor paga la tarifa de transferencia de datos entre regiones de la región productora a la región de consumo.

# Permisos que puede conceder a los recursos compartidos de datos

Puede concederles distintos tipos de objetos y varios permisos en un contexto de uso compartido de datos.

Bases de datos:

- CREATE
- USAGE
- ALTER
- DROP

Esquemas:

- CREATE
- USAGE
- ALTER
- DROP

Tablas:

- SELECT
- INSERT
- UPDATE
- DELETE
- TRUNCATE
- DROP
- ALTER
- REFERENCES

Funciones:

- EXECUTE

# Instrucciones SQL compatibles con escrituras mediante uso compartido de datos en consumidores

Se admiten las siguientes instrucciones del lenguaje de definición de datos (DDL) para el uso compartido de datos con escrituras:

- ALTER TABLE RENAME TO
- ALTER TABLE RENAME COLUMN TO
- ALTER TABLE ADD/DROP COLUMN
- ALTER SCHEMA RENAME TO
- { CREATE | DROP } SCHEMA
- { CREATE | DROP | SHOW } TABLE
- CREATE TABLE table_name AS
- BEGIN | START TRANSACTION
- END | COMMIT | ROLLBACK
- TRUNCATE

Se admiten las siguientes instrucciones del lenguaje de manipulación de datos (DML) para el uso compartido de datos con escrituras:

- SELECT
- INSERT
- INSERT INTO SELECT
- UPDATE
- DELETE
- MERGE
- COPY sin COMPUPDATE

Las siguientes instrucciones de análisis son compatibles con el uso compartido de datos con escrituras:

- ANALYZE. El consumidor ejecuta USER ANALYZE y envía las estadísticas al productor.

- Análisis activado por CTAS/COPY que se ejecuta en el consumidor. Esto incluye la multiplexación para consumidores pequeños.
- Análisis automático ejecutado en el productor después de COPY.

Las siguientes instrucciones de permisos son compatibles con el uso compartido de datos con escrituras:

- { GRANT | REVOKE } privilege_name ON OBJECT_TYPE object_name TO consumer_user
- SHOW GRANTS. Para obtener más información, consulte SHOW GRANTS.

## Instrucciones SQL no compatibles con escrituras mediante uso compartido de datos en consumidores

Lo siguiente no es compatible:

- Consultas de varias instrucciones a los almacenes de los consumidores cuando se escribe a los productores.
- Consultas de varias instrucciones a almacenamientos de consumidor en una base de datos diferente, si el comando anterior es una instrucción de lectura.
- Referencias a objetos que no son notaciones de tres puntos, como las notaciones de un punto o dos puntos, si no están conectadas a una base de datos compartida.
- Consultas de escalación simultáneas que se escriben de los consumidores a los productores.
- Trabajos de copia automática que se escriben de los consumidores a los productores.
- Trabajos de transmisión que se escriben de los consumidores a los productores.
- Consumidores que crean tablas de integración sin ETL en clústeres de productores. Para obtener información sobre las integraciones sin ETL, consulte Uso de integraciones sin ETL.
- Escribir en una tabla con una clave de clasificación intercalada a través de un recurso compartido de datos.
- Escribir en un procedimiento almacenado a través de un recurso compartido de datos.
- Escribir en una función definida por el usuario (UDF) de SQL a través de un recurso compartido de datos. Incluye UDF anidadas, de Python y Lambda.
- Instrucciones UPDATE, INSERTE o COPY en columnas de identidad para almacenamientos de consumidor con más sectores de computación que el productor.

- Instrucciones MERGE en tablas externas que no son de RMS para almacenamientos de consumidores, cuando se escribe en los productores.
- Instrucciones CREATE TABLE con:
  - Expresión DEFAULT establecida en el tipo de datos VARBYTE. El tipo de datos VARBYTE no se puede convertir de forma implícita en otros tipos de datos. Para obtener más información, consulte Función CAST.
  - Instrucciones AS OF SELECT con el parámetro NULL para almacenamientos de consumidores, al escribir en los productores.
  - Parámetro LIKE para almacenamientos de consumidores cuando se escribe en los productores.

# Regiones de AWS en las que está disponible el uso compartido de datos

En la siguiente tabla, se muestra la disponibilidad de las capacidades de lectura y escritura del uso compartido de datos.

| Región | Uso compartido de datos en la misma región | Uso compartido de datos entre regiones | Recursos compartidos de datos gobernados por AWS Lake Formation |
|---|---|---|---|
| Este de EE. UU. (Norte de Virginia) (us-east-1) | Sí | Sí | Sí |
| Este de EE. UU. (Ohio) (us-east-2) | Sí | Sí | Sí |
| EE. UU. Oeste (Norte de California) (us-west-1) | Sí | Sí | Sí |
| Oeste de EE. UU. (Oregón) (us-west-2) | Sí | Sí | Sí |

| Región | Uso compartido de datos en la misma región | Uso compartido de datos entre regiones | Recursos compartidos de datos gobernados por AWS Lake Formation |
|---|---|---|---|
| Asia-Pacífico (Hong Kong) (ap-east-1) | Sí | No | No |
| Asia Pacífico (Bombay) (ap-south-1) | Sí | Sí | Sí |
| Asia Pacífico (Hyderabad) (ap-south-2) | Sí | No | No |
| Asia-Pacífico (Tokio) (ap-northeast-1) | Sí | Sí | Sí |
| Asia-Pacífico (Singapur) (ap-southeast-1) | Sí | Sí | Sí |
| Asia-Pacífico (Sídney) (ap-southeast-2) | Sí | Sí | Sí |
| Asia-Pacífico (Yakarta); (ap-southeast-3) | Sí | No | No |
| Asia Pacífico (Melbourne) (ap-southeast-4) | Sí | No | No |
| Asia-Pacífico (Seúl) (ap-northeast-2) | Sí | Sí | Sí |
| Asia Pacific (Osaka) (ap-northeast-3) | Sí | Sí | Sí |

| Región | Uso compartido de datos en la misma región | Uso compartido de datos entre regiones | Recursos compartidos de datos gobernados por AWS Lake Formation |
|---|---|---|---|
| China (Pekín) (cn-north-1) | Sí | Sí | No |
| China (Ningxia) (cn-northwest-1) | Sí | Sí | No |
| África (Ciudad del Cabo) (af-south-1) | Sí | No | No |
| Oeste de Canadá (Calgary) (ca-west-1) | Sí | No | No |
| Canadá (centro) (ca-central-1) | Sí | Sí | Sí |
| Europa (Fráncfort) (eu-central-1) | Sí | Sí | Sí |
| Europa (Zúrich) (eu-central-2) | Sí | No | No |
| Europa (Irlanda) (eu-west-1) | Sí | Sí | Sí |
| Europa (Londres) (eu-west-2) | Sí | Sí | Sí |
| Europa (París) (eu-west-3) | Sí | Sí | Sí |
| UE (Milán) (eu-south-1) | Sí | No | No |

| Región | Uso compartido de datos en la misma región | Uso compartido de datos entre regiones | Recursos compartidos de datos gobernados por AWS Lake Formation |
|---|---|---|---|
| Europa (España) (eu-south-2) | Sí | No | No |
| Europa (Estocolmo) (eu-north-1) | Sí | Sí | Sí |
| Medio Oriente (EAU) (me-central-1) | Sí | No | No |
| Medio Oriente (Baréin) (me-south-1) | Sí | No | No |
| Israel (Tel Aviv) (il-central-1) | Sí | No | No |
| América del Sur (São Paulo) (sa-east-1) | Sí | Sí | Sí |
| AWS GovCloud (EE. UU. Este) (us-gov-east-1) | Sí | No | Sí |
| AWS GovCloud (EE. UU. Oeste) (us-gov-west-1) | Sí | No | Sí |

# Introducción al uso compartido de datos en Amazon Redshift

Puede utilizar una de las guías de esta sección para empezar a utilizar el uso compartido de datos.

Temas

- Introducción al uso compartido de datos de solo lectura en la consola
- Introducción al uso compartido de datos de solo lectura con la interfaz de SQL

- Introducción a las escrituras en varios almacenamientos mediante el uso compartido de datos en Amazon Redshift
- Introducción al uso compartido de datos con AWS CloudFormation en Amazon Redshift

## Introducción al uso compartido de datos de solo lectura en la consola

Con Amazon Redshift, puede administrar el uso compartido de datos mediante escrituras con la consola para controlar el acceso y gestionar los datos en todos los clústeres de Amazon Redshift y cuentas de AWS. En las siguientes secciones se proporcionan instrucciones paso a paso sobre configuración y administración de uso compartido de datos con escrituras mediante la consola de Amazon Redshift.

Utilice la consola de Amazon Redshift para administrar los datashares creados en su cuenta o compartidos por otras cuentas.

Necesita permisos para crear, editar o eliminar recursos compartidos de datos. Para obtener más información, consulta Administración de permisos para un recurso compartido de datos en Amazon Redshift.

- Si es el administrador de un productor, puede crear recursos compartidos de datos, añadir consumidores de datos, añadir objetos de recursos compartidos de datos, crear bases de datos a partir de recursos compartidos de datos, editar recursos compartidos de datos o eliminarlos desde la pestaña CLÚSTERES.

  En el menú de navegación, diríjase a la pestaña Clusters (Clústeres) y elija un clúster de la lista correspondiente. A continuación, realice alguna de las operaciones siguientes:

  - Elija la pestaña Datashares (Recursos compartidos de datos) y un recurso compartido de datos de la sección Datashares created in my namespace (Recursos compartidos de datos creados en mi espacio de nombres). A continuación, realice alguna de las operaciones siguientes:
    - Creación de un recurso compartido de datos

      Cuando se crea un datashare, puede agregar objetos de datashares o consumidores de datos. Para obtener más información, consulte Adición de objetos de recursos compartidos de datos a recursos compartidos de datos y Adición de consumidores de datos a recursos compartidos de datos.
    - Edición de datashares creados en su cuenta
    - Eliminación de un recurso compartido de datos creado en su cuenta

- Elija Datashares y seleccione un datashare de la sección Datashares from other clusters (Datashares de otros clústeres). A continuación, realice alguna de las operaciones siguientes:
  - Creación de un recurso compartido de datos
  - Creación de bases de datos a partir de datashares
- Seleccione Databases (Bases de datos) y elija una base de datos de la sección Databases (Bases de datos). Luego, elija Create datashare (Crear datashare). Para obtener más información, consulte Creación de bases de datos a partir de datashares.

> **Note**
>
> Para ver las bases de datos y los objetos que se encuentran dentro de ellas, o para ver los datashares del clúster, conéctese a una base de datos. Para obtener más información, consulte Conexión a una base de datos.

## Conexión a una base de datos

Con Amazon Redshift, puede establecer una conexión con el clúster de almacenamiento de datos y ejecutar consultas SQL, cargar datos o realizar tareas administrativas. La conexión a una base de datos se refiere al proceso de creación de un canal seguro entre una aplicación o herramienta cliente y el clúster de Amazon Redshift. En las siguientes secciones se proporcionarán instrucciones paso sobre cómo conectarse a una base de datos de Amazon Redshift.

Conéctese a una base de datos para ver las bases de datos y los objetos que se encuentran dentro de ellas en este clúster, o bien, para ver los datashares.

Las credenciales de usuario utilizadas para conectarse a una base de datos especificada deben tener los permisos necesarios para ver todos los recursos compartidos de datos.

Si no hay ninguna conexión local, realice alguna de las siguientes operaciones:

- En la página de detalles del clúster, en la pestaña Databases (Bases de datos), en la sección Databases (Bases de datos) o Datashare objects (Objetos de recurso compartido de datos), elija Connect to database (Conectarse a base de datos) para ver los objetos de base de datos del clúster.
- En la página de detalles del clúster, en la pestaña Datashares, realice alguna de las siguientes operaciones:

  - En la sección Datashares from other clusters (Datashares de otros clústeres), elija Connect to database (Conectarse a la base de datos) para ver los datashares de otros clústeres.
  - En la sección Datashares created in my cluster (Datashares creados en mi clúster), elija Connect to database (Conectarse a la base de datos) para ver los datashares de su clúster.
- En la ventana Connect to database (Conectarse a la base de datos), realice alguna de las siguientes operaciones:
  - Si elige Create a new connection (Crear una nueva conexión), elija AWS Secrets Manager para usar un secreto almacenado con objeto de autenticar el acceso a la conexión.

    O bien, elija Temporary credentials (Credenciales temporales) para utilizar credenciales de base de datos y con ellas autenticar el acceso de la conexión. Especifique valores para Database name (Nombre de base de datos) y Database user (Usuario de la base de datos).

    Elija Conectar.
  - Seleccione Use a recent connection (Utilizar una conexión reciente) para conectarse a otra base de datos para la que tenga los permisos necesarios.

    Amazon Redshift establece la conexión automáticamente.

Una vez establecida la conexión con la base de datos, puede comenzar a crear recursos compartidos de datos, consultarlos o crear bases de datos a partir de ellos.

## Creación de recursos compartidos de datos

Con Amazon Redshift, puede compartir datos en directo entre los clústeres de Amazon Redshift o las cuentas de AWS mediante recursos compartidos de datos. Un recurso compartido de datos es un objeto de consumidor-productor que le permite compartir datos en directo del clúster de Amazon Redshift con otros clústeres o cuentas de AWS. La creación de recursos compartidos de datos permite asegurar el uso compartido de datos mientras mantiene el control sobre el acceso y garantiza que los datos permanezcan actualizados. En las siguientes secciones se proporcionan detalles sobre la creación de recursos compartidos de datos y la agregación de objetos de base de datos, como esquemas, tablas y vistas, para compartir datos en directo de forma segura.

### Creación de un recurso compartido de datos

Un recurso compartido de datos es un contenedor lógico de objetos, permisos y consumidores de la base de datos. Los consumidores son clústeres aprovisionados de Amazon Redshift o espacios de nombres de Amazon Redshift sin servidor de su cuenta y de otras Cuentas de AWS. Cada

recurso compartido de datos está asociado a la base de datos en la que se creó y solo se pueden agregar objetos de esa base de datos. Como administrador de productores, puede crear recursos compartidos de datos en la consola y mediante SQL utilizando uno de los siguientes procedimientos.

Console

En la consola, puede crear recursos compartidos de datos desde las pestañas Recursos compartidos de datos de la página de detalles del clúster o del espacio de nombres. Una vez creado el recurso compartido de datos, puede crear bases de datos a partir de él en un consumidor como administrador de consumidores.

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. En el menú de navegación, elija Clusters (Clústeres) y, a continuación, elija su clúster. Se abrirá la página de detalles del clúster.
3. En la página de detalles del clúster o del espacio de nombres, en la pestaña Recursos compartidos de datos, en la sección Recursos compartidos de datos, conéctese a una base de datos si aún no está conectado. En la sección Recursos compartidos de datos creados en mi cuenta, elija Crear un recurso compartido de datos. Se abrirá la página Crear un recurso compartido de datos.
4. Elija Create datashare (Crear datashare). Solo puede crear un datashare a partir de una base de datos local. Si no se ha conectado a la base de datos, aparecerá la página Conectarse a la base de datos. Siga los pasos de Conexión a una base de datos para conectarse a una base de datos. Si hay una conexión reciente, se abrirá la página Crear un recurso compartido de datos.
5. En la sección Datashare information (Información sobre los recursos compartidos de datos), elija una de las siguientes opciones:
   - Elija Recurso compartido de datos para crear recursos compartidos de datos con el objetivo de compartir datos con fines de lectura o escritura entre diferentes almacenamientos de datos de Amazon Redshift (clústeres aprovisionados o puntos de conexión sin servidor) o bien en la misma Cuenta de AWS o en diferentes Cuentas de AWS.
   - Elija AWS Data Exchange datashare (Recurso compartido de datos de ) para crear recursos compartidos de datos con objeto de obtener licencias para los datos a través de AWS Data Exchange.
6. Especifique valores para Datashare name (Nombre del recurso compartido de datos), Database name (Nombre de la base de datos) y Publicly accessible (De acceso público).

Cuando cambie el nombre de la base de datos, establezca una nueva conexión a la base de datos.

7. Añada objetos a su recurso compartido de datos mediante las secciones Permisos con ámbito o Permisos directos. Para añadir objetos a un recurso compartido de datos, consulte Creación de un recurso compartido de datos en Amazon Redshift.
8. En la sección Consumidores de datos, puede optar por publicar en Amazon Redshift o en el AWS Glue Data Catalog, lo que inicia el proceso para compartir datos con Lake Formation. Al publicar el recurso compartido de datos en Amazon Redshift, los datos se comparten con otro espacio de nombres o cuenta de Amazon Redshift que actúa como consumidor.

   > **Note**
   >
   > Una vez creado el recurso compartido de datos, no puede editar la configuración para realizar la publicación en la otra opción.

9. Elija Create datashare (Crear datashare).

SQL

El siguiente comando crea un recurso compartido de datos:

```
CREATE DATASHARE salesshare;
```

En el momento de crear el recurso compartido de datos, cada recurso compartido de datos se asocia a una base de datos. Solo los objetos de esa base de datos se pueden compartir en ese datashare. Se pueden crear varios datashares en la misma base de datos con el mismo nivel de detalle de los objetos o uno diferente. No hay límite en cuanto a la cantidad de datashares que puede crear un clúster. También puede utilizar la consola de Amazon Redshift para crear datashares. Para obtener más información, consulte CREATE DATASHARE.

También puede controlar las restricciones de seguridad del recurso compartido de datos durante su creación. En el siguiente ejemplo, se muestra que el consumidor que cuenta con acceso de IP pública tiene permitido leer el recurso compartido de datos.

```
CREATE DATASHARE my_datashare [PUBLICACCESSIBLE = TRUE];
```

Si se establece PUBLICACCESSIBLE como TRUE, los consumidores pueden consultar su recurso compartido de datos desde clústeres de acceso público y grupos de trabajo aprovisionados. Omita esta opción o defínala explícitamente como falsa si no quiere permitirla.

Puede modificar las propiedades referidas al tipo de consumidores después de crear un recurso compartido de datos. Por ejemplo, puede definir que los clústeres que deseen consumir datos de un determinado datashare no puedan ser de acceso público. Las consultas de los clústeres consumidores que no cumplan las restricciones de seguridad especificadas en el recurso compartido de datos se rechazarán en el tiempo de ejecución de la consulta. Para obtener más información, consulte ALTER DATASHARE.

### Adición de objetos de recursos compartidos de datos a recursos compartidos de datos

Puede añadir objetos de base de datos de varios tipos en la consola y mediante SQL utilizando uno de los siguientes procedimientos.

#### Console

Puede añadir objetos al recurso compartido de datos utilizando las secciones Permisos con ámbito o Permisos directos. Seleccione Conceder permisos con ámbito o Conceder permisos directos para añadir objetos. Seleccione el botón Añadir para añadir objetos. Aparecerá un cuadro de diálogo. Siga estos pasos:

1. Si selecciona Conceder permisos con ámbito, aparece la página Conceder permisos con ámbito, donde puede conceder permisos con ámbito en la base de datos o el esquema. Los recursos compartidos de datos con permisos con ámbito tienen los permisos especificados en todos los objetos actuales y futuros de la base de datos o del esquema. Para obtener más información, consulte Permisos acotados.

    1. A continuación, seleccione Permisos con ámbito de base de datos para conceder permisos con ámbito en la base de datos. Al conceder permisos con ámbito, estos se aplican a la base de datos actual al crear el recurso compartido de datos. Estos permisos no se pueden conceder a objetos individuales y se aplican tanto a los objetos existentes como a los nuevos (esquemas, tablas, vistas y UDF).
    2. Seleccione los permisos con ámbito para los esquemas, tablas, vistas o funciones definidas por el usuario. Esto significa que todos los objetos de la base de datos tienen los permisos seleccionados que se han concedido a los consumidores. Seleccione Conceder para completar la concesión de los permisos con ámbito de la base de datos.

3. A continuación, seleccione Permisos con ámbito de esquema para conceder permisos con ámbito en el esquema. Al conceder permisos con ámbito de esquema, todos los objetos que se añaden al esquema tienen los permisos de recurso compartido de datos especificados.
4. Seleccione en el menú desplegable el esquema que desea añadir al recurso compartido de datos. Solo puede seleccionar un esquema a la vez. A continuación, seleccione los permisos directos que desee conceder en el esquema seleccionado.
5. Seleccione permisos con ámbito para sus objetos de esquema, como tablas, vistas y funciones definidas por el usuario. Los permisos se conceden a todos los objetos correspondientes del esquema. Pueden ser objetos existentes o que se añadan en el futuro. Cuando se aplica, el permiso no se puede eliminar de un objeto sin revocar los permisos con ámbito.
6. Seleccione Conceder para completar la concesión de los permisos con ámbito de esquema.

2. Si selecciona Conceder permisos directos, aparece la página Conceder permisos directos, en la que puede conceder permisos directos a cada objeto, como un esquema, una tabla, una vista o una función definida por el usuario. Para conceder permisos directos, primero debe añadir los esquemas pertinentes al recurso compartido de datos.

   1. A continuación, seleccione Conceder permisos directos a los esquemas para aplicar los permisos directos a un esquema específico. Luego, seleccione los permisos de esquema para los objetos de esquema, como tablas, vistas y funciones definidas por el usuario, y seleccione el esquema que desee añadir al recurso compartido de datos. Seleccione Conceder para completar la adición del esquema al recurso compartido de datos.
   2. Una vez que haya añadido un esquema al recurso compartido de datos, puede continuar añadiendo permisos directos para los objetos del esquema. Vuelva a seleccionar Conceder permisos directos. Aparece la página Conceder permisos directos. A continuación, vaya a las pestañas de permisos directos de los objetos del esquema.
   3. Seleccione Conceder permisos directos a tablas y vistas para conceder permisos directos de objeto en estos objetos. Seleccione los permisos directos necesarios y los objetos necesarios en la lista. Puede utilizar el campo de búsqueda para buscar objetos de recursos compartidos de datos. Seleccione Conceder para terminar de añadir tablas y vistas al recurso compartido de datos.
   4. Seleccione Conceder permisos directos a funciones definidas por el usuario para conceder permisos directos de objeto a funciones definidas por el usuario. Seleccione los permisos

directos necesarios y los objetos necesarios en la lista. Puede utilizar el campo de búsqueda para buscar objetos de recursos compartidos de datos. Seleccione Conceder para terminar de añadir funciones definidas por el usuario al recurso compartido de datos.

3. También puede elegir si desea Añadir objetos futuros. Si decide incluir los objetos compartidos de datos agregados al esquema, significa que los objetos agregados al esquema se agregan al recurso compartido de datos automáticamente.
4. Seleccione Agregar para completar la sección y añadir los objetos. Estos se enumeran bajo Objetos del recurso compartido de datos.
5. Tras añadir objetos, puede seleccionar objetos individuales y editar sus permisos. Si selecciona un esquema, aparece un cuadro de diálogo en el que se le pregunta si desea añadir Permisos acotados. Esto hace que cada objeto existente o añadido al esquema tenga un conjunto de permisos preseleccionados, adecuados para el tipo de objeto. Por ejemplo, el administrador puede configurar que todas las tablas añadidas tengan los permisos SELECT y UPDATE.
6. Todos los objetos de recursos compartidos de datos se muestran en las secciones Permisos con ámbito o Permisos directos.
7. En la sección Consumidores de datos, puede agregar espacios de nombres o cuentas de AWS como consumidores del recurso compartido de datos.
8. Elija Crear un recurso compartido de datos para guardar los cambios.

Después de crear el recurso compartido de datos, este aparece en la lista Recursos compartidos de datos creados en mi espacio de nombres. Si elige un recurso compartido de datos de la lista, puede ver sus consumidores, sus objetos y otras propiedades.

SQL

Con SQL, el propietario del recurso compartido de datos debe conceder USAGE en los esquemas que quiera añadir al recurso compartido de datos. GRANT se utiliza para conceder varias acciones en el esquema, incluidas CREATE y USAGE. Los esquemas contienen objetos compartidos:

```
CREATE SCHEMA myshared_schema1;
CREATE SCHEMA myshared_schema2;

GRANT USAGE ON SCHEMA myshared_schema1 TO DATASHARE my_datashare;
GRANT CREATE, USAGE ON SCHEMA myshared_schema2 TO DATASHARE my_datashare;
```

Como alternativa, el administrador puede seguir ejecutando los comandos ALTER para añadir un esquema al recurso compartido de datos. Solo se conceden permisos de USAGE cuando se agrega un esquema de esta manera.

```
ALTER DATASHARE my_datashare ADD SCHEMA myshared_schema1;
```

Una vez que el administrador ha agregado los esquemas, puede conceder permisos de recurso compartido de datos sobre los objetos del esquema. Pueden ser permisos de lectura y escritura. En el ejemplo GRANT ALL se muestra cómo conceder todos los permisos.

```
GRANT SELECT, INSERT ON TABLE myshared_schema1.table1, myshared_schema1.table2,
 myshared_schema2.table1
TO DATASHARE my_datashare;

GRANT ALL ON TABLE myshared_schema1.table4 TO DATASHARE my_datashare;
```

Puede seguir ejecutando comandos como ALTER DATASHARE para añadir tablas. Cuando lo haga, solo se concederán los permisos SELECT a los objetos añadidos.

```
ALTER DATASHARE my_datashare ADD TABLE myshared_schema1.table1,
 myshared_schema1.table2, myshared_schema2.table1;
```

Puede conceder permisos con ámbito a un recurso compartido de datos en todos los objetos de un tipo en una base de datos o esquema. Los recursos compartidos de datos con permisos con ámbito tienen los permisos especificados en todos los objetos actuales y futuros de la base de datos o del esquema.

Puede ver el alcance de los permisos limitados en el nivel de base de datos en SVV_DATABASE_PRIVILEGES. Puede ver el alcance de los permisos limitados en el nivel de esquema en SVV_SCHEMA_PRIVILEGES.

Esta es la sintaxis para conceder permisos con ámbito a recursos compartidos de datos. Para obtener más información sobre los permisos acotados, consulte Permisos acotados.

```
GRANT { CREATE | USAGE | ALTER | DROP } [,...] | ALL [ PRIVILEGES ] }FOR SCHEMAS IN
DATABASE db_name
TO DATASHARE { datashare_name}

GRANT { { SELECT | INSERT | UPDATE | DELETE | DROP | ALTER | TRUNCATE | REFERENCES }
 [, ...] } | ALL [PRIVILEGES] } }FOR TABLES IN
```

```
{SCHEMA schema_name [DATABASE db_name ] | DATABASE db_name }
TO DATASHARE { datashare_name}

GRANT { EXECUTE | ALL [ PRIVILEGES ] }FOR FUNCTIONS IN
{SCHEMA schema_name [DATABASE db_name ] | DATABASE db_name }
TO DATASHARE { datashare_name}
```

Adición de consumidores de datos a recursos compartidos de datos

Puede agregar uno o más consumidores de datos a los datashares. Los consumidores de datos pueden ser espacios de nombres que identifican de forma exclusiva los clústeres de Amazon Redshift o las Cuentas de AWS.

Debe elegir explícitamente si desea habilitar o no la opción de compartir su recurso compartido de datos con los clústeres de acceso público.

- Seleccione Agregar espacios de nombres al recurso compartido de datos. El espacio de nombres es el identificador global único (GUID) del clúster de Amazon Redshift.
- Elija Add Cuentas de AWS (Agregar Cuentas de AWS) al recurso compartido de datos. Las Cuentas de AWS especificadas deben tener permisos de acceso al recurso compartido de datos.

## Autorización o retiro de la autorización de los recursos compartidos de datos

Como administrador de productores, elija a qué consumidores de datos autorizará para que accedan a los recursos compartidos de datos o a qué consumidores de datos retirará la autorización. Los consumidores de datos autorizados reciben notificaciones para tomar medidas en los datashares. Si va a añadir un espacio de nombres como consumidor de datos, no tiene que realizar el proceso de autorización.

Requisito previo: para autorizar el datashare o retirar la autorización, debe haber al menos un consumidor de datos agregado al datashare.

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. En el menú de navegación, elija Datashares (Recursos compartidos de datos). Se abrirá la página con la lista de datashares.
3. Elija In my account (En mi cuenta).

4. En la sección Datashares in my account (Datashares en mi cuenta), realice alguna de las siguientes operaciones:
   - Elija uno o más recursos compartidos de datos que desee autorizar. Se abrirá la página Authorize data consumers (Autorizar consumidores de datos). A continuación elija Authorize.

     Si ha elegido Publicar en AWS Glue Data Catalog al crear el recurso compartido de datos, solo puede conceder la autorización del recurso compartido de datos a una cuenta de Lake Formation.

     Para el recurso compartido de datos de AWS Data Exchange, solo puede autorizar un recurso compartido de datos a la vez.

     Cuando autoriza un recuso compartido de datos de AWS Data Exchange, comparte el recurso compartido de datos con el servicio AWS Data Exchange y permite que AWS Data Exchange administre el acceso al recurso compartido de datos en su nombre. Para que los consumidores obtengan acceso, AWS Data Exchange agrega cuentas consumidoras como consumidores de datos al recurso compartido de datos de AWS Data Exchange cuando se suscriben a los productos. AWS Data Exchange no tiene acceso de lectura al recurso compartido de datos.
   - Elija uno o más clústeres consumidores a los que desee retirar la autorización. A continuación, elija Remove authorization (Retirar autorización).

Una vez autorizados los consumidores de datos, ellos pueden acceder a los objetos de los datashares y crear una base de datos consumidora para consultar los datos.

Después de retirar la autorización, los consumidores de datos pierden de inmediato el acceso a los recursos compartidos de datos.

## Administración de recursos compartidos de datos de otras cuentas como consumidor

### Eliminación de la asociación del recurso compartido de datos a los consumidores de datos

Como administrador de consumidores, puede eliminar la asociación de los recursos compartidos de datos de los consumidores de datos.

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. En el menú de navegación, elija Datashares (Recursos compartidos de datos). Se abrirá la página con la lista de datashares.

3. Elija From other accounts (De otras cuentas).
4. En la sección Datashares from other accounts (Datashares de otras cuentas), elija el datashare cuya asociación a los consumidores de datos desea eliminar.
5. En la sección Data consumers (Consumidores en datos), elija uno o más consumidores de datos para los cuales desea eliminar la asociación. A continuación, elija Remove association (Eliminar asociación).
6. Cuando aparezca la página de eliminación de la asociación, elija Remove association (Eliminar asociación).

Después de eliminar la asociación, los consumidores de datos perderán el acceso al datashare. Puede cambiar la asociación de consumidores de datos en cualquier momento.

Rechazo de los recursos compartidos de datos

Como administrador de consumidores, puede rechazar cualquier recurso compartido de datos cuyo estado sea disponible o activo. Después de rechazar un recurso compartido de datos, los usuarios del clúster consumidor perderán el acceso a dicho recurso. Amazon Redshift no devuelve el recurso compartido de datos rechazado si se llama a la operación de la API `DescribeDataSharesForConsumer`. Si el administrador de productores ejecuta la operación de la API `DescribeDataSharesForProducer`, verá que se rechazó el recurso compartido de datos. Una vez que se rechaza un recurso compartido de datos, el administrador de productores puede volver a autorizar el recurso compartido de datos a un clúster de consumidor y el administrador del clúster de consumidores puede decidir asociar su cuenta de AWS al recurso compartido de datos o rechazarlo.

Si su cuenta de AWS tiene una asociación a un recurso compartido de datos y una asociación pendiente a un recurso compartido de datos administrado por Lake Formation, al rechazar la asociación del recurso compartido de datos que administra Lake Formation también se rechaza el recurso compartido de datos original. Para rechazar una asociación específica, el administrador de productores puede eliminar la autorización de un recurso compartido de datos especificado. Esta acción no afecta a otros recursos compartidos de datos.

Para rechazar un recurso compartido de datos, utilice la consola de AWS, la operación de la API `RejectDataShare` o `reject-datashare` en la AWS CLI.

Para rechazar un recurso compartido de datos mediante la consola de AWS:

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. En el menú de navegación, elija Recursos compartidos de datos.
3. Elija From other accounts (De otras cuentas).
4. En la sección Datashares from other accounts (Datashares de otras cuentas), elija el datashare que desee rechazar. Cuando aparezca la página Decline datashare (Rechazar recurso compartido de datos), elija Decline (Rechazar).

Después de rechazar los datashares, no puede revertir el cambio. Amazon Redshift elimina los recursos compartidos de datos de la lista. Para volver a ver el recurso compartido de datos, el administrador del productor debe volver a autorizarlo.

## Administración de recursos compartidos de datos existentes

Con Amazon Redshift, puede administrar los recursos compartidos de datos existentes para controlar el acceso a los datos en un clúster de Amazon Redshift. En las siguientes secciones se proporcionan instrucciones paso sobre la administración de recursos compartidos de datos en el entorno de Amazon Redshift.

### Visualización de los datashares

Vea los datashares desde la pestaña DATASHARES o la pestaña CLUSTERS (CLÚSTERES).

- Use la pestaña DATASHARES para ver una lista de los datashares de su cuenta o de otras cuentas.
  - Para ver los datashares que se han creado en su cuenta, elija In my account (En mi cuenta) y, luego, seleccione el datashare que desee consultar.
  - Para ver los datashares que se comparten desde otras cuentas, elija From other accounts (De otras cuentas) y, luego, seleccione los datashares que desee consultar.
- Use la pestaña CLUSTERS (CLÚSTERES) para enumerar los datashares de su clúster o de otros clústeres.

  Conéctese a una base de datos. Para obtener más información, consulte Conexión a una base de datos.

A continuación, elija un datashare de la sección Datashares from other clusters (Datashares de otros clústeres) o la sección Datashares created in my cluster (Datashares creados en mi clúster) para ver sus detalles.

### Eliminación de objetos de los datashares

Puede eliminar uno o más objetos de un recurso compartido de datos mediante el siguiente procedimiento.

Elimine uno o más objetos de un recurso compartido de datos.

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. En el menú de navegación, elija Clusters (Clústeres) y, a continuación, elija su clúster. Se abrirá la página de detalles del clúster.
3. Seleccione Datashares.
4. En la sección Datashares created in my account (Datashares creados en mi cuenta), elija Connect to database (Conectarse a la base de datos). Para obtener más información, consulte Conexión a una base de datos.
5. Elija el datashare que desee modificar y, a continuación, seleccione Edit (Editar). Se abrirá la página de detalles del datashare.
6. Para quitar uno o más objetos del datashare, realice alguna de las siguientes operaciones:

   - Para quitar esquemas del datashare, elija uno o más esquemas. Luego, seleccione Remove (Quitar). Amazon Redshift quitará los esquemas especificados y todos los objetos de los esquemas especificados del datashare.
   - Para quitar tablas y vistas del datashare, elija una o más tablas y vistas. Luego, seleccione Remove (Quitar). También puede elegir Remove by schema (Quitar por esquema) para quitar todas las tablas y las vistas de los esquemas especificados.
   - Para quitar funciones definidas por el usuario del datashare, elija una o más funciones definidas por el usuario. Luego, seleccione Remove (Quitar). También puede elegir Remove by schema (Quitar por esquema) para quitar todas las funciones definidas por el usuario de los esquemas especificados.

Eliminación de consumidores de datos de los datashares

Puede quitar uno o más consumidores de datos de un datashare. Los consumidores de datos pueden ser espacios de nombres que identifican de forma exclusiva los clústeres de Amazon Redshift o las cuentas de AWS.

Elija uno o más consumidores de datos en los ID de espacios de nombres o la cuenta de AWS y, a continuación, seleccione Eliminar.

Amazon Redshift quita los consumidores de datos especificados del datashare. Los consumidores especificados perderán de inmediato el acceso al datashare.

Edición de datashares creados en su cuenta

Edite los datashares creados en su cuenta con la consola. Conéctese primero a una base de datos para ver la lista de datashares creados en su cuenta.

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. En el menú de navegación, elija Clusters (Clústeres) y, a continuación, elija su clúster. Se abrirá la página de detalles del clúster.
3. Seleccione Datashares.
4. En la sección Datashares created in my account (Datashares creados en mi cuenta), elija Connect to database (Conectarse a la base de datos). Para obtener más información, consulte Conexión a una base de datos.
5. Elija el datashare que desee modificar y, a continuación, seleccione Edit (Editar). Se abrirá la página de detalles del datashare.
6. Realice todos los cambios en la sección Datashare objects (Objetos de datashares) o en la sección Data consumers (Consumidores de datos).

   > **Note**
   >
   > Si eligió publicar su recurso compartido de datos en AWS Glue Data Catalog, no podrá editar la configuración para publicarlo en otras cuentas de Amazon Redshift.

7. Seleccione Save changes (Guardar cambios).

Amazon Redshift actualizará sus datashares con los cambios.

Eliminación de un recurso compartido de datos creado en su cuenta

Elimine los datashares creados en su cuenta con la consola. Conéctese primero a una base de datos para ver la lista de datashares creados en su cuenta.

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. En el menú de navegación, elija Clusters (Clústeres) y, a continuación, elija su clúster. Se abrirá la página de detalles del clúster.
3. Seleccione Datashares. Aparecerá la lista de datashares.
4. En la sección Datashares created in my account (Datashares creados en mi cuenta), elija Connect to database (Conectarse a la base de datos). Para obtener más información, consulte Conexión a una base de datos.
5. Elija uno o más datashares que desee eliminar y, a continuación, seleccione Delete (Eliminar). Se abrirá la página Delete datashares (Eliminar datashares).

   La eliminación de un recurso compartido de datos con Lake Formation no elimina automáticamente los permisos asociados en Lake Formation. Para eliminarlos, vaya a la consola de Lake Formation.
6. Escriba Delete (Eliminar) para confirmar la eliminación de los datashares especificados.
7. Elija Eliminar.

Después de eliminar los datashares, sus consumidores perderán el acceso a ellos.

## Consultas de recursos compartidos de datos

Con Amazon Redshift, puede consultar datos entre recursos compartidos de datos de clústeres de productores para acceder de forma segura a los datos en directo sin necesidad de copiarlos ni transferirlos. En las siguientes secciones se describen los detalles de la configuración y consulta de los recursos compartidos de datos en el entorno de Amazon Redshift.

Creación de bases de datos a partir de datashares

Para comenzar a consultar los datos de un datashare, cree una base de datos a partir de él. Solo puede crear una base de datos a partir de un datashare especificado.

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.

2. En el menú de navegación, elija Clusters (Clústeres) y, a continuación, elija su clúster. Se abrirá la página de detalles del clúster.
3. Seleccione Datashares. Aparecerá la lista de datashares.
4. En la sección Datashares from other clusters (Datashares de otros clústeres), elija Connect to database (Conectarse a la base de datos). Para obtener más información, consulte Conexión a una base de datos.
5. Seleccione el datashare a partir del cual desea crear bases de datos y, a continuación, elija Create database from datashare (Crear base de datos a partir del datashare). Se abrirá la página Create database from datashare (Crear base de datos a partir del datashare).
6. En Database name (Nombre de base de datos), especifique un nombre para la base de datos. El nombre de la base de datos debe tener entre 1 y 64 caracteres alfanuméricos (solo letras minúsculas) y no puede ser una palabra reservada.
7. Seleccione Create  (Crear).

Después de crear la base de datos, puede consultar los datos que se encuentran en ella.

## Administración de recursos compartidos de datos de AWS Data Exchange

Con Amazon Redshift, puede compartir y recibir datos en directo de forma segura de AWS Data Exchange sin tener que crear ni administrar extracciones o canalizaciones de datos. La administración de recursos compartidos de datos de AWS Data Exchange le permite suscribirse a productos de datos de terceros e integrar flujos de datos en directo directamente en el almacenamiento de datos de Amazon Redshift. En las siguientes secciones, se muestra la administración de recursos compartidos de datos de AWS Data Exchange dentro de los clústeres de Amazon Redshift.

### Creación de conjuntos de datos en AWS Data Exchange

Cree conjuntos de datos en AWS Data Exchange.

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. En el menú de navegación, elija Clusters (Clústeres) y, a continuación, elija su clúster. Se abrirá la página de detalles del clúster.
3. Seleccione Datashares (Recursos compartidos de datos).
4. En la sección Datashares created in my account (Recursos compartidos de datos creados en mi cuenta), elija un recurso compartido de datos de AWS Data Exchange.

5. Elija Create data set on AWS Data Exchange (Crear conjunto de datos en AWS Data Exchange). Para obtener más información, consulte Publicación de un producto nuevo.

### Edición de los conjuntos compartidos de datos en AWS Data Exchange

Edite los recursos compartidos de datos de AWS Data Exchange con la consola. Conéctese primero a una base de datos para ver la lista de datashares creados en su cuenta.

En los recursos compartidos de datos de AWS Data Exchange, no puede realizar cambios en los consumidores de datos.

Para editar la configuración de acceso público para los recursos compartidos de datos de AWS Data Exchange, utilice el editor de consultas v2. Amazon Redshift genera un valor aleatorio por única vez para configurar la variable de sesión y permitir que se desactive esta configuración. Para obtener más información, consulte Notas de uso de ALTER DATASHARE.

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. En el menú de navegación, elija Clusters (Clústeres) y, a continuación, elija su clúster. Se abrirá la página de detalles del clúster.
3. En el menú del navegador, elija Editor (Editor) y luego Query editor v2 (Editor de consultas v2).
4. Si es la primera vez que utiliza el editor de consultas v2, configure su Cuenta de AWS. De manera predeterminada, se utiliza una clave propia de AWS para cifrar los recursos. Para obtener más información sobre cómo configurar la Cuenta de AWS, consulte Configuración de la Cuenta de AWS en la Guía de administración de Amazon Redshift.
5. Para conectarse al clúster en el que se encuentra su recurso compartido de datos de AWS Data Exchange, elija Database (Base de datos) y el nombre del clúster en el panel de vista de árbol. Si se le solicita, ingrese los parámetros de conexión.
6. Copie la siguiente instrucción SQL. En el siguiente ejemplo, se cambia la configuración de acceso público del recurso compartido de datos SalesShare.

```
ALTER DATASHARE salesshare SET PUBLICACCESSIBLE FALSE;
```

7. Para ejecutar la instrucción SQL copiada, elija Queries (Consultas) y pegue la instrucción SQL en el área de consultas. A continuación, elija Run (Ejecutar).

   Aparecerá el siguiente error:

```
ALTER DATASHARE salesshare SET PUBLICACCESSIBLE FALSE;
ERROR:  Alter of ADX-managed datashare salesshare requires session variable
 datashare_break_glass_session_var to be set to value 'c670ba4db22f4b'
```

El valor “c670ba4db22f4b” es un valor aleatorio único que Amazon Redshift genera cuando se produce una operación no recomendada.

8. Copie y pegue la siguiente instrucción de muestra en el área de consultas. Luego, ejecute el comando. El comando `SET datashare_break_glass_session_var` aplica un permiso para permitir una operación no recomendada para un recurso compartido de datos de AWS Data Exchange.

```
SET datashare_break_glass_session_var to 'c670ba4db22f4b';
```

9. Vuelva a ejecutar la instrucción ALTER DATASHARE.

```
ALTER DATASHARE salesshare;
```

Amazon Redshift actualizará sus datashares con los cambios.

Eliminación de los recursos compartidos de datos de AWS Data Exchange creados en su cuenta

Elimine los recursos compartidos de datos de AWS Data Exchange creados en su cuenta con la consola. Conéctese primero a una base de datos para ver la lista de datashares creados en su cuenta.

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. En el menú de navegación, elija Clusters (Clústeres) y, a continuación, elija su clúster. Se abrirá la página de detalles del clúster.
3. En el menú del navegador, elija Editor (Editor) y luego Query editor v2 (Editor de consultas v2).
4. Si es la primera vez que utiliza el editor de consultas v2, configure su Cuenta de AWS. De manera predeterminada, se utiliza una clave propia de AWS para cifrar los recursos. Para obtener más información sobre cómo configurar la Cuenta de AWS, consulte Configuración de la Cuenta de AWS en la Guía de administración de Amazon Redshift.

5. Para conectarse al clúster en el que se encuentra su recurso compartido de datos de AWS Data Exchange, elija Database (Base de datos) y el nombre del clúster en el panel de vista de árbol. Si se le solicita, ingrese los parámetros de conexión.
6. Copie la siguiente instrucción SQL. En el siguiente ejemplo, se elimina el recurso compartido de datos SalesShare.

```
DROP DATASHARE salesshare
```

7. Para ejecutar la instrucción SQL copiada, elija Queries (Consultas) y pegue la instrucción SQL en el área de consultas. A continuación, elija Run (Ejecutar).

   Aparecerá el siguiente error:

```
ERROR:  Drop of ADX-managed datashare salesshare requires session variable
 datashare_break_glass_session_var to be set to value '620c871f890c49'
```

   El valor “620c871f890c49” es un valor aleatorio único que Amazon Redshift genera cuando se produce una operación no recomendada.
8. Copie y pegue la siguiente instrucción de muestra en el área de consultas. Luego, ejecute el comando. El comando `SET datashare_break_glass_session_var` aplica un permiso para permitir una operación no recomendada para un recurso compartido de datos de AWS Data Exchange.

```
SET datashare_break_glass_session_var to '620c871f890c49';
```

9. Vuelva a ejecutar la instrucción DROP DATASHARE.

```
DROP DATASHARE salesshare;
```

Después de eliminar los recursos compartidos de datos, sus consumidores perderán el acceso a ellos.

Eliminar un recurso compartido de datos de AWS Data Exchange compartido puede infringir los términos del producto de datos en AWS Data Exchange.

# Introducción al uso compartido de datos de solo lectura con la interfaz de SQL

Con Amazon Redshift, puede compartir datos de forma segura en clústeres de Amazon Redshift, lo que habilita a los consumidores de datos para consultar y acceder en directo a los datos sin copiarlos o replicarlos. El uso compartido de datos le permite crear y configurar recursos compartidos de datos, que son objetos del productor que hacen referencia a los objetos de la base de datos que desea compartir.

Puede compartir datos con fines de lectura entre distintos clústeres de Amazon Redshift dentro o entre Cuentas de AWS, o entre Regiones de AWS.

Temas


- Uso compartido de acceso de lectura a los datos en una Cuenta de AWS
- Uso compartido de datos con vistas en Amazon Redshift
- Agregación de tablas de lago de datos a un recurso compartido de datos
- Compartir datos entre Cuentas de AWS
- Compartir datos entre Regiones de AWS
- Uso compartido de datos de Amazon Redshift con licencia en AWS Data Exchange
- Introducción a los recursos compartidos de datos administrados por AWS Lake Formation


## Uso compartido de acceso de lectura a los datos en una Cuenta de AWS

Con Amazon Redshift, puede compartir el acceso de lectura a los datos entre distintos usuarios o grupos de bases de datos dentro de la misma cuenta de AWS. Esta característica le permite controlar los privilegios de acceso a los datos de forma detallada, lo que garantiza que solo los usuarios o grupos autorizados puedan leer conjuntos de datos específicos.

Uso compartido de datos con fines de lectura como administrador de productores o como propietario de la base de datos

1. Cree datashares en su clúster. Para obtener más información, consulte CREATE DATASHARE.

```
CREATE DATASHARE salesshare;
```

Los superusuarios de clústeres y los propietarios de bases de datos pueden crear datashares. Cada datashare se asocia a una base de datos durante la creación. Solo los objetos de esa base de datos se pueden compartir en ese datashare. Se pueden crear varios datashares en la misma base de datos con el mismo nivel de detalle de los objetos o uno diferente. No hay límite en cuanto a la cantidad de datashares que puede crear un clúster.

También puede utilizar la consola de Amazon Redshift para crear datashares. Para obtener más información, consulte Creación de un recurso compartido de datos.

2. Delegue permisos para operar en el datashare. Para obtener más información, consulte GRANT o REVOKE.

   En el siguiente ejemplo, se otorgan permisos a `dbuser` para `salesshare`.

   ```
   GRANT ALTER, SHARE ON DATASHARE salesshare TO dbuser;
   ```

   Los superusuarios del clúster y los propietarios del recurso compartido de datos pueden otorgar o revocar permisos de modificación sobre este a usuarios adicionales.

3. Agregue objetos a los datashares o quítelos. Para agregar objetos a un datashare, agregue el esquema antes de agregar los objetos. Cuando se agrega un esquema, Amazon Redshift no agrega todos los objetos que contiene. Asegúrese de agregarlos explícitamente. Para obtener más información, consulte ALTER DATASHARE.

   ```
   ALTER DATASHARE salesshare ADD SCHEMA PUBLIC;
   ALTER DATASHARE salesshare ADD TABLE public.tickit_sales_redshift;
   ALTER DATASHARE salesshare ADD ALL TABLES IN SCHEMA PUBLIC;
   ```

   También se pueden agregar vistas a un datashare.

   ```
   CREATE VIEW public.sales_data_summary_view AS SELECT * FROM
    public.tickit_sales_redshift;
   ALTER DATASHARE salesshare ADD VIEW public.sales_data_summary_view;
   ```

   Utilice ALTER DATASHARE para compartir esquemas, tablas, vistas y funciones de un esquema determinado. Los superusuarios, los propietarios de recursos compartidos de datos o los usuarios que tienen permisos ALTER u ALL sobre el recurso compartido de datos pueden modificarlo para agregarle o quitarle objetos. Los usuarios deben tener los permisos para poder

agregar objetos al recurso compartido de datos o quitarlos. Los usuarios también deben ser los propietarios de los objetos o disponer de permisos SELECT, USAGE u ALL sobre ellos.

También puede utilizar GRANT para agregar objetos al recurso compartido de datos. En este ejemplo, puede ver cómo:

```
GRANT SELECT ON TABLE public.tickit_sales_redshift TO DATASHARE salesshare;
```

La funcionalidad de esta sintaxis es equivalente a `ALTER DATASHARE salesshare ADD TABLE public.tickit_sales_redshift;`.

Utilice la cláusula INCLUDENEW para agregar al datashare nuevas tablas, vistas o funciones definidas por el usuario (UDF) de SQL que se crearon en un esquema especificado. Solo los superusuarios pueden modificar esta propiedad para cada par de esquema y datashare.

```
ALTER DATASHARE salesshare ADD SCHEMA PUBLIC;
ALTER DATASHARE salesshare SET INCLUDENEW = TRUE FOR SCHEMA PUBLIC;
```

También puede utilizar la consola de Amazon Redshift para agregar objetos a los datashares o quitarlos. Para obtener más información, consulte Adición de objetos de recursos compartidos de datos a recursos compartidos de datos, Eliminación de objetos de los datashares y Edición de datashares creados en su cuenta.

4. Agregue consumidores a los datashares o quítelos. En el siguiente ejemplo, se añade el espacio de nombres de consumidor a `salesshare`. El espacio de nombres es el identificador único de carácter general (GUID) del espacio de nombres del clúster consumidor de la cuenta. Para obtener más información, consulte GRANT o REVOKE.

```
GRANT USAGE ON DATASHARE salesshare TO NAMESPACE
 '13b8833d-17c6-4f16-8fe4-1a018f5ed00d';
```

Solo se pueden otorgar permisos a un consumidor de recurso compartido de datos en una instrucción GRANT.

Los superusuarios de clústeres y los propietarios de objetos del recurso compartido de datos o los usuarios que tienen el permiso SHARE sobre el recurso compartido de datos pueden agregarle consumidores o quitárselos. Para ello, utilizan las instrucciones GRANT USAGE o REVOKE USAGE.

Para encontrar el espacio de nombres del clúster que ve en este momento, puede utilizar el comando SELECT CURRENT_NAMESPACE. Para encontrar el espacio de nombres de un clúster diferente dentro de la misma Cuenta de AWS, vaya a la página de detalles del clúster en la consola de Amazon Redshift. En esa página, busque el campo de espacio de nombres que se acaba de agregar.

También puede utilizar la consola de Amazon Redshift para agregar o quitar consumidores de datos en los recursos compartidos de datos. Para obtener más información, consulte Adición de consumidores de datos a recursos compartidos de datos y Eliminación de consumidores de datos de los datashares.

5. (Opcional) Agregue restricciones de seguridad al datashare. En el siguiente ejemplo, se muestra que el clúster consumidor que cuenta con acceso de IP pública tiene permitido leer el datashare. Para obtener más información, consulte ALTER DATASHARE.

```
ALTER DATASHARE salesshare SET PUBLICACCESSIBLE = TRUE;
```

Puede modificar las propiedades referidas al tipo de consumidores después de la creación del datashare. Por ejemplo, puede definir que los clústeres que deseen consumir datos de un determinado datashare no puedan ser de acceso público. Las consultas de los clústeres consumidores que no cumplan las restricciones de seguridad especificadas en el recurso compartido de datos se rechazarán en el tiempo de ejecución de la consulta.

También puede utilizar la consola de Amazon Redshift para modificar datashares. Para obtener más información, consulte Edición de datashares creados en su cuenta.

6. Enumere los datashares que se hayan creado en el clúster y examine su contenido.

En el siguiente ejemplo, se muestra la información de un recurso compartido de datos denominado `salesshare`.

```
DESC DATASHARE salesshare;

 producer_account  |          producer_namespace          | share_type | share_name
 | object_type |           object_name          |   include_new
-------------------+--------------------------------------+------------
+------------+-------------+---------------------------------+-------------------
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | table       | public.tickit_users_redshift   |
```

```
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | table       | public.tickit_venue_redshift   |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | table       | public.tickit_category_redshift|
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | table       | public.tickit_date_redshift    |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | table       | public.tickit_event_redshift   |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | table       | public.tickit_listing_redshift |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | table       | public.tickit_sales_redshift   |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | schema      | public                         |  t
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | view        | public.sales_data_summary_view |
```

En el siguiente ejemplo, se muestran los datashares salientes de un clúster productor.

```
SHOW DATASHARES LIKE 'sales%';
```

El resultado tiene un aspecto similar al siguiente.

```
share_name | share_owner  | source_database | consumer_database | share_type |
 createdate      | is_publicaccessible  | share_acl | producer_account |
 producer_namespace
-----------+-------------+-----------------+-------------------+------------
+---------------------+----------------------+-----------+-----------------
+--------------------------------------
salesshare |    100       | dev             |                   |  OUTBOUND
 | 2020-12-09 02:27:08 |          True        |           |   123456789012   |
 13b8833d-17c6-4f16-8fe4-1a018f5ed00d
```

Para obtener más información, consulte DESC DATASHARE y SHOW DATASHARES.

También puede utilizar SVV_DATASHARES, SVV_DATASHARE_CONSUMERS y SVV_DATASHARE_OBJECTS para ver los datashares, los objetos dentro del datashare y sus consumidores.

7. Elimine datashares. Para obtener más información, consulte DROP DATASHARE.

Puede eliminar los objetos del datashare en cualquier momento con DROP DATASHARE. Los superusuarios del clúster y los propietarios de los datashares pueden eliminarlos.

En el siguiente ejemplo, se elimina un recurso compartido de datos denominado `salesshare`.

```
DROP DATASHARE salesshare;
```

También puede utilizar la consola de Amazon Redshift para eliminar los datashares. Para obtener más información, consulte Eliminación de un recurso compartido de datos creado en su cuenta.

8. Utilice ALTER DATASHARE para quitar objetos de los datashares en cualquier parte de este. Utilice REVOKE USAGE ON para revocar los permisos sobre el datashare otorgados a determinados consumidores. Revoca los permisos USAGE de los objetos que se encuentran dentro de un recurso compartido de datos y desactiva de forma instantánea el acceso para todos los clústeres consumidores. Las listas de los datashares y las consultas de metadatos, como las listas de bases de datos y tablas, no devuelven los objetos compartidos después de haber revocado el acceso.

   ```
   ALTER DATASHARE salesshare REMOVE TABLE public.tickit_sales_redshift;
   ```

   También puede utilizar la consola de Amazon Redshift para modificar datashares. Para obtener más información, consulte Edición de datashares creados en su cuenta.

9. Revoque el acceso a los datashares desde los espacios de nombres si ya no desea compartir los datos con los consumidores.

   ```
   REVOKE USAGE ON DATASHARE salesshare FROM NAMESPACE
    '13b8833d-17c6-4f16-8fe4-1a018f5ed00d';
   ```

   También puede utilizar la consola de Amazon Redshift para modificar datashares. Para obtener más información, consulte Edición de datashares creados en su cuenta.

Uso compartido de datos con fines de lectura como administrador de consumidores

1. Enumere los datashares que se ponen a su disposición y vea su contenido. Para obtener más información, consulte DESC DATASHARE y SHOW DATASHARES.

En el siguiente ejemplo, se muestra la información de los datashares entrantes de un espacio de nombres productor específico. Cuando ejecuta DESC DATASHARE como administrador de consumidores, debe especificar la opción NAMESPACE para ver los recursos compartidos de datos entrantes.

```
DESC DATASHARE salesshare OF NAMESPACE '13b8833d-17c6-4f16-8fe4-1a018f5ed00d';


 producer_account  |          producer_namespace          | share_type | share_name
 | object_type |           object_name           |   include_new
-------------------+--------------------------------------+------------
+------------+-------------+---------------------------------+------------------
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    | salesshare
 | table       | public.tickit_users_redshift    |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    | salesshare
 | table       | public.tickit_venue_redshift    |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    | salesshare
 | table       | public.tickit_category_redshift |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    | salesshare
 | table       | public.tickit_date_redshift     |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    | salesshare
 | table       | public.tickit_event_redshift    |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    | salesshare
 | table       | public.tickit_listing_redshift  |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    | salesshare
 | table       | public.tickit_sales_redshift    |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    | salesshare
 | schema      | public                          |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    | salesshare
 | view        | public.sales_data_summary_view  |
```

Solo los superusuarios del clúster pueden hacer esto. También puede utilizar SVV_DATASHARES para ver los datashares y SVV_DATASHARE_OBJECTS para ver los objetos que se encuentran dentro del datashare.

En el siguiente ejemplo, se muestran los datashares entrantes de un clúster consumidor.

```
SHOW DATASHARES LIKE 'sales%';
```

```
 share_name | share_owner | source_database | consumer_database | share_type
 | createdate | is_publicaccessible | share_acl | producer_account |
 producer_namespace
-----------+-------------+-----------------+-------------------+-----------
+-----------+---------------------+-----------+------------------
+--------------------------------------
 salesshare |             |                 |                   | INBOUND
    |            |         t           |           |   123456789012   |
 13b8833d-17c6-4f16-8fe4-1a018f5ed00d
```

2. Como superusuario de bases de datos, puede crear bases de datos locales que hacen referencia a los recursos compartidos de datos. Para obtener más información, consulte CREATE DATABASE.

```
CREATE DATABASE sales_db FROM DATASHARE salesshare OF NAMESPACE
 '13b8833d-17c6-4f16-8fe4-1a018f5ed00d';
```

   Si desea tener un control más detallado del acceso a los objetos de la base de datos local, utilice la cláusula WITH PERMISSIONS al crear la base de datos. De este modo, podrá conceder permisos en el nivel de objeto para los objetos de la base de datos en el paso 4.

```
CREATE DATABASE sales_db WITH PERMISSIONS FROM DATASHARE salesshare OF NAMESPACE
 '13b8833d-17c6-4f16-8fe4-1a018f5ed00d';
```

   Puede ver las bases de datos que creó a partir del datashare consultando la vista SVV_REDSHIFT_DATABASES. Puede conectarse a estas bases de datos directamente o puede conectarse a una base de datos local en el clúster del consumidor y realizar una consulta entre bases de datos para consultar los datos de las bases de datos de recursos compartidos de datos. No puede crear un datashare encima de los objetos de base de datos creados a partir de un datashare existente. No obstante, puede copiar los datos en una tabla diferente en el clúster consumidor, llevar a cabo el procesamiento que sea necesario y, a continuación, compartir los objetos nuevos que se hayan creado.

   También puede utilizar la consola de Amazon Redshift para crear bases de datos a partir de datashares. Para obtener más información, consulte Creación de bases de datos a partir de datashares.

3. (Opcional) Cree esquemas externos para referenciar esquemas específicos de la base de datos consumidora importados en el clúster consumidor y para asignarles permisos pormenorizados. Para obtener más información, consulte CREATE EXTERNAL SCHEMA.

```
CREATE EXTERNAL SCHEMA sales_schema FROM REDSHIFT DATABASE 'sales_db' SCHEMA
 'public';
```

4. Conceda permisos en las bases de datos y las referencias de esquema creadas a partir de los recursos compartidos de datos a usuarios y roles del clúster del consumidor según sea necesario. Para obtener más información, consulte GRANT o REVOKE.

```
GRANT USAGE ON DATABASE sales_db TO Bob;
```

```
GRANT USAGE ON SCHEMA sales_schema TO ROLE Analyst_role;
```

Si ha creado su base de datos sin WITH PERMISSIONS, solo podrá asignar permisos a toda la base de datos creada a partir del recurso compartido de datos a los usuarios y los roles. En algunos casos, necesita controles pormenorizados sobre un subconjunto de los objetos de base de datos creados a partir del datashare. Si es así, puede crear una referencia de esquema externa que apunte a esquemas específicos del datashare (como se describe en el paso anterior) y proporcionar permisos pormenorizados en el nivel del esquema.

También puede crear vistas de enlace de tiempo de ejecución encima de los objetos compartidos y utilizarlas para asignar permisos pormenorizados. También puede considerar la posibilidad de indicar a los clústeres productores que creen datashares adicionales en su nombre con el nivel de detalle requerido.

Si ha creado la base de datos con WITH PERMISSIONS en el paso 2, debe asignar permisos de nivel de objeto a los objetos de la base de datos compartida. Un usuario que solo tenga el permiso USAGE no podrá acceder a ningún objeto de una base de datos creada con WITH PERMISSIONS hasta que se le concedan permisos adicionales en el nivel de objeto.

```
GRANT SELECT ON sales_db.public.tickit_sales_redshift to Bob;
```

5. Consulte los datos de los objetos compartidos que se encuentran en los datashares.

Los usuarios y los roles que tengan permisos en las bases de datos del consumidor y los esquemas de los clústeres del consumidor pueden explorar y navegar por los metadatos de cualquier objeto compartido. También pueden navegar por los objetos locales de un clúster consumidor y explorarlos. Para ello, utilizan controladores JDBC u ODBC, o vistas SVV_ALL y SVV_REDSHIFT.

Los clústeres productores pueden tener muchos esquemas en la base de datos y muchas tablas y vistas dentro de cada esquema. Los usuarios del lado del consumidor solo pueden ver el subconjunto de objetos que se han puesto a disposición a través del datashare. Estos usuarios no pueden ver todos los metadatos del clúster productor. Este enfoque ayuda a proporcionar un control pormenorizado de la seguridad de los metadatos con el uso compartido de datos.

Continúa conectándose a las bases de datos locales del clúster. No obstante, ahora, también es posible leer de las bases de datos y los esquemas que se crean a partir de los datashares utilizando la notación de tres partes database.schema.table. Puede realizar consultas que abarquen todas las bases de datos visibles para usted. Estas pueden ser bases de datos locales del clúster o bases de datos creadas a partir de los datashares. Los clústeres consumidores no pueden conectarse a las bases de datos que se hayan creado a partir de los datashares.

Puede acceder a los datos mediante la cualificación completa. Para obtener más información, consulte Ejemplos de consultas entre bases de datos.

```
SELECT * FROM sales_db.public.tickit_sales_redshift ORDER BY 1,2 LIMIT 5;

 salesid | listid | sellerid | buyerid | eventid | dateid | qtysold | pricepaid |
 commission |      saletime
---------+--------+----------+---------+---------+--------+---------+-----------
+------------+---------------------
       1 |      1 |    36861 |   21191 |    7872 |   1875 |       4 |    728.00 |
   109.20 | 2008-02-18 02:36:48
       2 |      4 |     8117 |   11498 |    4337 |   1983 |       2 |     76.00 |
    11.40 | 2008-06-06 05:00:16
       3 |      5 |     1616 |   17433 |    8647 |   1983 |       2 |    350.00 |
    52.50 | 2008-06-06 08:26:17
       4 |      5 |     1616 |   19715 |    8647 |   1986 |       1 |    175.00 |
    26.25 | 2008-06-09 08:38:52
       5 |      6 |    47402 |   14115 |    8240 |   2069 |       2 |    154.00 |
    23.10 | 2008-08-31 09:17:02
```

Solo se pueden utilizar las instrucciones SELECT en los objetos compartidos. No obstante, puede crear tablas en el clúster consumidor consultando los datos de los objetos compartidos de una base de datos local diferente.

Además de las consultas, los consumidores pueden crear vistas en los objetos compartidos. Solo se admiten las vistas de enlace de tiempo de ejecución o las vistas materializadas.

Amazon Redshift no admite las vistas normales en los datos compartidos. Las vistas que crean los consumidores pueden abarcar varias bases de datos locales o bases de datos creadas a partir de datashares. Para obtener más información, consulte CREATE VIEW.

```
// Connect to a local cluster database

// Create a view on shared objects and access it.
CREATE VIEW sales_data
AS SELECT *
FROM sales_db.public.tickit_sales_redshift
WITH NO SCHEMA BINDING;

SELECT * FROM sales_data;
```

## Uso compartido de datos con vistas en Amazon Redshift

El clúster productor puede compartir vistas normales, de enlace de tiempo de ejecución y materializadas. Cuando comparte vistas normales o de enlace de tiempo de ejecución, no tiene que compartir las tablas base. En la siguiente tabla, se muestra cómo se admiten las vistas en el uso compartido de datos.

| Nombre de la vista | ¿Se puede agregar esta vista a un datashare? | ¿Un consumidor puede crear esta vista en los objetos de los datashares en distintos clústeres? |
|---|---|---|
| Vista normal | Sí | No |
| Vista de enlace en tiempo de ejecución | Sí | Sí |
| Vista materializada | Sí | Sí, pero solo con una actualización completa |

La siguiente consulta muestra el resultado de una vista normal que es compatible con el uso compartido de datos. Para obtener información acerca de la definición de la vista normal, consulte CREATE VIEW.

```
SELECT * FROM tickit_db.public.myevent_regular_vw
ORDER BY eventid LIMIT 5;

   eventid  |  eventname
  ----------+-------------
     3835   | LeAnn Rimes
     3967   | LeAnn Rimes
     4856   | LeAnn Rimes
     4948   | LeAnn Rimes
     5131   | LeAnn Rimes
```

La siguiente consulta muestra el resultado de una vista de enlace en tiempo de ejecución que es compatible con el uso compartido de datos. Para obtener información sobre la definición de la vista de enlace en tiempo de ejecución, consulte CREATE VIEW.

```
SELECT * FROM tickit_db.public.event_lbv
ORDER BY eventid LIMIT 5;

 eventid | venueid | catid | dateid |          eventname           |      starttime
 --------+---------+-------+--------+------------------------------
+---------------------
     1   |   305   |   8   |  1851  |        Gotterdammerung       | 2008-01-25
 14:30:00
     2   |   306   |   8   |  2114  |           Boris Godunov      | 2008-10-15
 20:00:00
     3   |   302   |   8   |  1935  |              Salome          | 2008-04-19
 14:30:00
     4   |   309   |   8   |  2090  |  La Cenerentola (Cinderella) | 2008-09-21
 14:30:00
     5   |   302   |   8   |  1982  |          Il Trovatore        | 2008-06-05
 19:00:00
```

La siguiente consulta muestra el resultado de una vista materializada que es compatible con el uso compartido de datos. Para obtener información acerca de la definición de la vista materializada, consulte CREATE MATERIALIZED VIEW.

```
SELECT * FROM tickit_db.public.tickets_mv;
```

```
 catgroup | qtysold
----------+---------
 Concerts | 195444
   Shows  | 149905
```

Puede mantener tablas comunes para todos los usuarios de un clúster productor. También puede compartir subconjuntos de datos filtrados por las columnas de dimensión, como `tenant_id` (`account_id` o `namespace_id`), con los clústeres consumidores. Para ello, puede definir una vista en la tabla base con un filtro en estas columnas de ID; por ejemplo, `current_aws_account = tenant_id`. En el lado del consumidor, cuando consulta la vista, solo ve las filas que califican para su cuenta. Para ello, puede utilizar las funciones de contexto `current_aws_account` y `current_namespace` de Amazon Redshift.

La siguiente consulta devuelve el ID de la cuenta en la que se encuentra el clúster actual de Amazon Redshift. Puede ejecutar esta consulta si está conectado a Amazon Redshift.

```
select current_user, current_aws_account;

current_user | current_aws_account
-------------+--------------------
dwuser       |    111111111111
(1row)
```

La siguiente consulta devuelve el espacio de nombres del clúster actual de Amazon Redshift. Puede ejecutar esta consulta si está conectado a la base de datos.

```
select current_user, current_namespace;

current_user | current_namespace
-------------+--------------------------------------
dwuser       | 86b5169f-01dc-4a6f-9fbb-e2e24359e9a8
(1 row)
```

Actualización incremental para obtener vistas materializadas en un recurso compartido de datos

Amazon Redshift admite la actualización incremental de las vistas materializadas en un recurso compartido de datos de consumidores cuando se comparten las tablas base. La actualización incremental es una operación en la que Amazon Redshift identifica los cambios en la tabla o tablas base que se produjeron después de la actualización anterior y actualiza solo los

registros correspondientes de la vista materializada. Para obtener más información sobre este comportamiento, consulte CREATE MATERIALIZED VIEW.

## Agregación de tablas de lago de datos a un recurso compartido de datos

Con un recurso compartido de datos, un productor de datos puede compartir de forma segura objetos de base de datos de gran detalle, como esquemas y tablas, con consumidores de la misma cuenta de AWS o de cuentas diferentes. El productor también puede compartir objetos entre regiones. En este tema se describe cómo agregar objetos de un lago de datos, específicamente, del catálogo de datos de AWS Glue, a un recurso compartido de datos. Incluye dos casos de uso:

- Agregación de una vista de enlace tardío a un recurso compartido de datos que haga referencia a una tabla de un lago de datos: esto resulta práctico para un consumidor, ya que es probable que la configuración preliminar, como la definición de los permisos en los datos de origen externos, por ejemplo, con Lake Formation, ya esté completada. Un beneficio adicional es que una vista agregada a un recurso compartido de datos puede unir tablas del lago de datos con tablas nativas de Redshift.
- Agregación de una tabla de un esquema externo a un recurso compartido de datos directamente: esto hace que los objetos del lago de datos estén disponibles para los consumidores sin capas ni lógica adicionales. Los consumidores pueden consultar la tabla o unirla a las tablas del consumidor.

Estos casos se aplican después de hacer referencia a una tabla del catálogo de datos de AWS de Redshift mediante CREATE EXTERNAL SCHEMA. Cualquier tabla del catálogo de datos de AWS puede ser el origen.

> **Note**
>
> Las tablas de lago de datos que se agregan a un recurso compartido de datos pueden incluir tablas registradas en Lake Formation y tablas de catálogos de datos de AWS Glue.

### Creación de un esquema externo y una tabla externa

Puede crear un esquema externo y una tabla externa para agregarlos al recurso compartido de datos en las secciones siguientes. Estos son pasos preliminares. Si ya lo ha hecho, puede omitir esta sección.

1. En el productor, cree un esquema externo que haga referencia a los datos del lago de datos almacenados en Amazon S3. El esquema externo hace referencia al catálogo de datos de AWS Glue. El rol y la región de la muestra son ejemplos:

```
CREATE EXTERNAL SCHEMA external_schema_name FROM DATA CATALOG
DATABASE 'glue_database_name'
IAM_ROLE 'arn:aws:iam::123456789012:role/sample-role'
REGION 'us-east-1';
```

2. Cree una tabla de lago de datos en el esquema externo.

```
CREATE EXTERNAL TABLE external_schema_name.sales(
salesid INTEGER,
sellerid INTEGER,
buyerid INTEGER,
saledate DATE,
pricepaid DECIMAL(8,2))
ROW FORMAT delimited
FIELDS TERMINATED BY '\t'
STORED AS textfile
LOCATION 's3://redshift-downloads/tickit/spectrum/sales/';
```

   El ejemplo incluye `LOCATION`. Debe estar en el formulario `s3://{bucket_name}/{folder}/`, el que se especifica la carpeta. La carpeta debe tener una longitud mínima de un carácter. Si lo desea, puede incluir subcarpetas. Para ver ejemplos adicionales de creación de tablas en un lago de datos, consulte Ejemplos de CREATE EXTERNAL TABLE.

   > **Note**
   >
   > Solo se permite el uso compartido para las tablas en las que el rol de IAM del productor tenga acceso SELECT sobre la tabla.

Agregue una vista de enlace tardío que haga referencia a una tabla de un lago de datos a un recurso compartido de datos

Cuando crea tablas en función de un esquema externo del catálogo de datos de AWS y desea agregarlas a un recurso compartido de datos, la forma más común de hacerlo es agregar una vista de enlace tardío de Redshift que haga referencia a la tabla que creó, que contiene datos del lago de datos. El procedimiento siguiente muestra los pasos:

1. Cree una vista de enlace tardío que haga referencia a la tabla externa que creó anteriormente:

```
CREATE VIEW lbv AS
select * from external_schema_name.sales, other_schema.t1
WITH NO SCHEMA BINDING;
```

2. Agregue el esquema de la vista al recurso compartido de datos. Este es el esquema local que contiene la vista de enlace tardío.

```
ALTER DATASHARE dsx_datashare ADD SCHEMA public;
```

3. Agregue el esquema que contiene la tabla a la que hace referencia la vista de enlace tardío al recurso compartido de datos. Es necesario agregar el esquema para cualquier tabla base a la que se haga referencia en una vista que se agregue a un recurso compartido de datos, ya sea que el esquema contenga objetos de una base de datos local u objetos de un lago de datos. Tenga en cuenta que debe agregar este esquema antes de agregar la vista de enlace tardío.

```
ALTER DATASHARE dsx_datashare ADD SCHEMA external_schema_name;
ALTER DATASHARE dsx_datashare ADD SCHEMA other_schema;
```

4. Agregue la vista al recurso compartido de datos, mediante un comando SQL. Tenga en cuenta que el nombre de la tabla incluye el prefijo del esquema.

```
ALTER DATASHARE my_datashare ADD TABLE public.lbv;
```

5. Confirme que la vista y los esquemas se hayan agregado correctamente al recurso compartido de datos:

```
SELECT * FROM svv_datashare_objects WHERE share_name = 'my_datashare';
```

6. El administrador consumidor crea una base de datos a partir del recurso compartido de datos y, a continuación, concede el uso a los usuarios consumidores.

Tras completar los pasos, los usuarios consumidores de la base de datos con acceso a la vista del recurso compartido de datos pueden consultar los datos.

Agregación de una tabla de lago de datos directamente a un recurso compartido de datos

La agregación de una tabla de un esquema externo a un recurso compartido de datos es similar a agregar una vista. Esto funciona bien en los casos en los que un consumidor quiere consultar la tabla

del lago de datos en su estado original o si quiere unirla a las tablas del almacenamiento de datos del consumidor. Los pasos siguientes muestran cómo agregar una tabla de lago de datos a un recurso compartido de datos mediante SQL.

1. Cree un esquema externo y una tabla externa, tal y como se describe en la primera sección de este tema.
2. Descubra las tablas existentes en el esquema externo para confirmar que la tabla que ha creado está disponible:

   ```
   SELECT * FROM svv_external_tables WHERE schemaname = 'external_schema_name';
   ```

3. Agregue el esquema externo al recurso compartido de datos:

   ```
   ALTER DATASHARE my_datashare ADD SCHEMA external_schema_name;
   ```

4. Agregue la tabla externa al recurso compartido de datos. Tenga en cuenta que el nombre de la tabla incluye el prefijo del esquema:

   ```
   ALTER DATASHARE my_datashare ADD TABLE external_schema_name.sales;
   ```

5. Confirme que la tabla se ha agregado correctamente al recurso compartido de datos:

   ```
   SELECT * FROM svv_datashare_objects WHERE share_name = 'my_datashare';
   ```

   Para obtener instrucciones más detalladas, consulte Compartir el acceso de lectura a los datos de una cuenta de AWS.
6. En el caso del consumidor, que es la base de datos que recibe los datos compartidos, el administrador asocia el recurso compartido de datos para que los usuarios puedan consultarlas en las tablas compartidas. Para obtener más información sobre cómo realizar este paso, consulte Administrar los datos compartidos de otras cuentas como consumidor.

Una vez que los administradores hayan completado los pasos, los usuarios de la base de datos del consumidor pueden escribir consultas para recuperar los datos de la tabla compartida y unirlos a otras tablas del consumidor.

### Notas de uso para agregar objetos de lago de datos a un recurso compartido de datos

Hay varios elementos que se deben tener en cuenta al utilizar tablas y vistas de un lago de datos en un recurso compartido de datos:

- Iniciar sesión con AWS CloudTrail: la cuenta del productor de datos puede usar los registros de AWS CloudTrail para auditar cuándo se accede a las tablas del lago de datos compartidas a través de un recurso compartido de datos:
  - Uso de datos de registro para controlar el acceso a los datos: los registros de CloudTrail registran detalles sobre quién accede a las tablas compartidas, incluidos los productores y consumidores de recurso compartido de datos de Redshift. Los identificadores están disponibles en el campo `ExternalId` en los registros de `AssumeRole` CloudTrail. El propietario de los datos puede configurar limitaciones adicionales al acceso a los datos en una política de IAM mediante acciones. Para obtener más información sobre cómo definir el acceso a los datos mediante políticas, consulte Acceso a cuentas de AWS propiedad de terceros.
- Seguridad y permisos de consumidor: en el caso de las tablas registradas de Lake Formation, Lake Formation protege los recursos de Amazon S3 y los pone a disposición mediante las credenciales proporcionadas por Lake Formation.

### Consideraciones de facturación para agregar objetos de lago de datos a un recurso compartido de datos

A continuación, se detalla cómo se atribuyen los costos del almacenamiento y el escaneo de los objetos del lago de datos en un recurso compartido de datos:

- Cuando un consumidor consulta objetos compartidos de un lago de datos, el costo del escaneo se factura al consumidor.
  - Cuando el consumidor es un clúster aprovisionado, Redshift usa Redshift Spectrum para escanear los datos de Amazon S3. Por lo tanto, el costo de Spectrum se factura a la cuenta del consumidor.
  - Cuando el consumidor es un grupo de trabajo de Amazon Redshift sin servidor, Spectrum no cobra de forma independiente.
- Los costes de almacenamiento y operaciones de Amazon S3, como la lista de los buckets, se facturan a la cuenta propietaria de cada bucket de Amazon S3.

Para obtener información adicional sobre facturación para Amazon Redshift sin servidor, consulte Facturación para Amazon Redshift sin servidor. Encontrará más información sobre facturación y precios en Precios de Amazon Redshift.

## Compartir datos entre Cuentas de AWS

Puede compartir datos con fines de lectura entre Cuentas de AWS. El uso compartido de datos entre Cuentas de AWS funciona de manera similar al uso compartido de datos dentro de una cuenta. La diferencia es que se requiere la autorización de ambas partes para compartir datos entre Cuentas de AWS. Los administradores de una cuenta productora pueden autorizar a las cuentas consumidoras para acceder a los datashares o elegir no autorizar ningún acceso. Para utilizar un recurso compartido de datos autorizado, el administrador de la cuenta consumidora puede asociar este recurso. El administrador puede asociar el recurso compartido de datos a la Cuenta de AWS completa o a clústeres específicos de la cuenta consumidora, o rechazar el recurso compartido de datos. Para obtener más información acerca del uso compartido de datos dentro de una cuenta, consulte Uso compartido de acceso de lectura a los datos en una Cuenta de AWS.

Un recurso compartido de datos puede tener consumidores de datos que sean espacios de nombres de la misma cuenta o de Cuentas de AWS diferentes. No es necesario crear datashares independientes para compartirlos dentro de una cuenta y entre cuentas.

Para compartir datos entre cuentas, tanto el clúster productor como el consumidor deben estar cifrados.

Cuando se comparten datos con Cuentas de AWS, los administradores de productores comparten los datos con la Cuenta de AWS como una entidad. El administrador de consumidores puede decidir qué espacios de nombres de la cuenta de consumidor pueden acceder a un recurso compartido de datos.

Temas

- Acciones del administrador de productores
- Acciones del administrador de una cuenta consumidora
- Acciones del administrador de consumidores

### Acciones del administrador de productores

Con Amazon Redshift, puede realizar tareas administrativas en los clústeres de productores para administrar la ingesta de datos y el procesamiento de cargas.

Si es administrador de productores o el propietario de una base de datos, siga estos pasos:

1. Cree datashares en el clúster y agregue objetos a los datashares. Para obtener instrucciones más detalladas sobre cómo crear datashares y agregarles objetos, consulte Uso compartido de acceso de lectura a los datos en una Cuenta de AWS. Para obtener más información acerca de CREATE DATASHARE y ALTER DATASHARE, consulte CREATE DATASHARE y ALTER DATASHARE.

   En el siguiente ejemplo, se agregan diferentes objetos de recursos compartidos de datos al recurso compartido de datos `salesshare`.

   ```
   -- Add schema to datashare
   ALTER DATASHARE salesshare ADD SCHEMA PUBLIC;

   -- Add table under schema to datashare
   ALTER DATASHARE salesshare ADD TABLE public.tickit_sales_redshift;

   -- Add view to datashare
   ALTER DATASHARE salesshare ADD TABLE public.sales_data_summary_view;

   -- Add all existing tables and views under schema to datashare (does not include
    future table)
   ALTER DATASHARE salesshare ADD ALL TABLES in schema public;
   ```

   También puede utilizar la consola de Amazon Redshift para crear o modificar los datashares. Para obtener más información, consulte Creación de un recurso compartido de datos y Edición de datashares creados en su cuenta.

2. Delegue permisos para operar en el datashare. Para obtener más información, consulte GRANT o REVOKE.

   En el siguiente ejemplo, se otorgan permisos a `dbuser` para `salesshare`.

   ```
   GRANT ALTER, SHARE ON DATASHARE salesshare TO dbuser;
   ```

   Los superusuarios del clúster y los propietarios del recurso compartido de datos pueden otorgar o revocar permisos de modificación sobre este a usuarios adicionales.

3. Agregue consumidores a los datashares o quítelos. En el siguiente ejemplo, se agrega el ID de la Cuenta de AWS a `salesshare`. Para obtener más información, consulte GRANT o REVOKE.

   ```
   GRANT USAGE ON DATASHARE salesshare TO ACCOUNT '123456789012';
   ```

Solo se pueden otorgar permisos a un consumidor de datos en una instrucción GRANT.

Los superusuarios de clústeres y los propietarios de objetos del recurso compartido de datos o los usuarios que tienen los permisos SHARE sobre el recurso compartido de datos pueden agregarle consumidores o quitárselos. Para ello, utilizan las instrucciones GRANT USAGE o REVOKE USAGE.

También puede utilizar la consola de Amazon Redshift para agregar o quitar consumidores de datos en los recursos compartidos de datos. Para obtener más información, consulte Adición de consumidores de datos a recursos compartidos de datos y Eliminación de consumidores de datos de los datashares.

4. (Opcional) Revoque el acceso al recurso compartido de datos desde las Cuentas de AWS si ya no desea compartir los datos con los consumidores.

```
REVOKE USAGE ON DATASHARE salesshare FROM ACCOUNT '123456789012';
```

Si es el administrador de una cuenta productora, siga estos pasos:

Después de otorgar el uso a Cuenta de AWS, el estado del recurso compartido de datos es `pending_authorization`. El administrador de la cuenta productora debe autorizar los datashares a través de la consola de Amazon Redshift y elegir los consumidores de datos.

Inicie sesión en https://console.aws.amazon.com/redshiftv2/. Luego elija a qué consumidores de datos autorizará para que accedan a los recursos compartidos de datos o a cuáles retirará la autorización. Los consumidores de datos autorizados reciben notificaciones para tomar medidas en los datashares. Si va a añadir un espacio de nombres como consumidor de datos, no tiene que realizar el proceso de autorización. Una vez autorizados los consumidores de datos, ellos pueden acceder a los objetos de los datashares y crear una base de datos consumidora para consultar los datos. Para obtener más información, consulte Autorización o retiro de la autorización de los recursos compartidos de datos.

Uso compartido de permisos de escritura de los datos entre cuentas

Con Amazon Redshift, puede compartir datos entre cuentas de AWS y conceder permisos de escritura, lo que permite la colaboración y el uso compartido de datos entre equipos u organizaciones. El uso compartido de datos entre cuentas le permite establecer una cuenta de proveedor de datos que crea y administra bases de datos, esquemas y tablas, que luego se

pueden compartir de forma segura con las cuentas de los consumidores de datos. En las siguientes secciones se muestra el proceso de configuración del uso compartido de datos entre cuentas y la concesión del acceso de escritura en Amazon Redshift.

Acciones del administrador de una cuenta consumidora

Con Amazon Redshift, puede administrar las cuentas de los consumidores y controlar su acceso a los recursos de almacenamiento de datos.

Si es el administrador de una cuenta consumidora, siga estos pasos:

Para asociar uno o más recursos compartidos de datos que se comparten desde otras cuentas con toda la Cuenta de AWS o con espacios de nombres específicos de su cuenta, utilice la consola de Amazon Redshift.

Inicie sesión en https://console.aws.amazon.com/redshiftv2/. Luego, asocie uno o más recursos compartidos de datos que se comparten desde otras cuentas con toda la Cuenta de AWS o con espacios de nombres específicos de su cuenta. Para obtener más información, consulte Asociación de un recurso compartido de datos desde otra Cuenta de AWS en Amazon Redshift.

Después de asociar la Cuenta de AWS o espacios de nombres específicos, los recursos compartidos de datos están a su disposición para que los consuma. También puede cambiar la asociación del datashare en cualquier momento. Cuando se cambia la asociación de espacios de nombres individuales a una Cuenta de AWS, Amazon Redshift sobrescribe los espacios de nombres con la información de la Cuenta de AWS. Cuando se cambia la asociación de una Cuenta de AWS a espacios de nombres específicos, Amazon Redshift sobrescribe la información de la Cuenta de AWS con la información del espacio de nombres. Todos los espacios de nombres de la cuenta obtienen acceso a los datos.

Acciones del administrador de consumidores

Con Amazon Redshift, puede realizar tareas administrativas en los clústeres de consumidores para administrar la ingesta de datos y el procesamiento de cargas.

Si es administrador de consumidores, siga estos pasos:

1. Enumere los datashares que se ponen a su disposición y consulte su contenido. El contenido de los recursos compartidos de datos solo está disponible cuando el administrador de productores ha autorizado los recursos compartidos de datos y el administrador de consumidores ha aceptado y asociado los recursos compartidos de datos. Para obtener más información, consulte DESC DATASHARE y SHOW DATASHARES.

En el siguiente ejemplo, se muestra la información de los datashares entrantes de un espacio de nombres productor específico. Cuando ejecuta DESC DATASHARE como administrador de consumidores, debe especificar el NAMESPACE y el ID de la cuenta para ver los recursos compartidos de datos entrantes. Para los recursos compartidos de datos salientes, especifique el nombre del recurso compartido de datos.

```
SHOW DATASHARES LIKE 'sales%';

share_name | share_owner | source_database | consumer_database | share_type
 | createdate | is_publicaccessible | share_acl | producer_account |
 producer_namespace
-----------+-------------+-----------------+-------------------+------------
+------------+---------------------+-----------+------------------
+--------------------------------------
salesshare |             |                 |                   | INBOUND    |
            |        t            |           | 123456789012    | 'dd8772e1-
d792-4fa4-996b-1870577efc0d'
```

```
DESC DATASHARE salesshare OF ACCOUNT '123456789012' NAMESPACE 'dd8772e1-
d792-4fa4-996b-1870577efc0d';


 producer_account |          producer_namespace          | share_type | share_name |
 object_type |           object_name
-----------------+--------------------------------------+------------+------------
+-------------+---------------------------------
 123456789012     | dd8772e1-d792-4fa4-996b-1870577efc0d | INBOUND    | salesshare |
 table       | public.tickit_users_redshift
 123456789012     | dd8772e1-d792-4fa4-996b-1870577efc0d | INBOUND    | salesshare |
 table       | public.tickit_venue_redshift
 123456789012     | dd8772e1-d792-4fa4-996b-1870577efc0d | INBOUND    | salesshare |
 table       | public.tickit_category_redshift
 123456789012     | dd8772e1-d792-4fa4-996b-1870577efc0d | INBOUND    | salesshare |
 table       | public.tickit_date_redshift
 123456789012     | dd8772e1-d792-4fa4-996b-1870577efc0d | INBOUND    | salesshare |
 table       | public.tickit_event_redshift
 123456789012     | dd8772e1-d792-4fa4-996b-1870577efc0d | INBOUND    | salesshare |
 table       | public.tickit_listing_redshift
 123456789012     | dd8772e1-d792-4fa4-996b-1870577efc0d | INBOUND    | salesshare |
 table       | public.tickit_sales_redshift
```

```
 123456789012     | dd8772e1-d792-4fa4-996b-1870577efc0d | INBOUND    | salesshare |
 schema      | public
(8 rows)
```

Solo los superusuarios del clúster pueden hacer esto. También puede utilizar SVV_DATASHARES para ver los datashares y SVV_DATASHARE_OBJECTS para ver los objetos que se encuentran dentro del datashare.

En el siguiente ejemplo, se muestran los datashares entrantes de un clúster consumidor.

```
SELECT * FROM SVV_DATASHARES WHERE share_name LIKE 'sales%';

share_name | share_owner | source_database | consumer_database | share_type
 | createdate | is_publicaccessible | share_acl | producer_account |
 producer_namespace
-----------+-------------+-----------------+-------------------+------------
+------------+---------------------+-----------+------------------
+--------------------------------------
salesshare |             |                 |                   | INBOUND    |
            |        t            |           | 123456789012      | 'dd8772e1-
d792-4fa4-996b-1870577efc0d'
```

```
SELECT * FROM SVV_DATASHARE_OBJECTS WHERE share_name LIKE 'sales%';
 share_type | share_name | object_type |           object_name           |
 producer_account |          producer_namespace
------------+------------+-------------+---------------------------------
+------------------+--------------------------------------
 INBOUND    | salesshare | table       | public.tickit_users_redshift    |
 123456789012     | dd8772e1-d792-4fa4-996b-1870577efc0d
 INBOUND    | salesshare | table       | public.tickit_venue_redshift    |
 123456789012     | dd8772e1-d792-4fa4-996b-1870577efc0d
 INBOUND    | salesshare | table       | public.tickit_category_redshift |
 123456789012     | dd8772e1-d792-4fa4-996b-1870577efc0d
 INBOUND    | salesshare | table       | public.tickit_date_redshift     |
 123456789012     | dd8772e1-d792-4fa4-996b-1870577efc0d
 INBOUND    | salesshare | table       | public.tickit_event_redshift    |
 123456789012     | dd8772e1-d792-4fa4-996b-1870577efc0d
 INBOUND    | salesshare | table       | public.tickit_listing_redshift  |
 123456789012     | dd8772e1-d792-4fa4-996b-1870577efc0d
 INBOUND    | salesshare | table       | public.tickit_sales_redshift    |
 123456789012     | dd8772e1-d792-4fa4-996b-1870577efc0d
```

```
 INBOUND    | salesshare | schema      | public                          |
 123456789012     | dd8772e1-d792-4fa4-996b-1870577efc0d
(8 rows)
```

2. Cree bases de datos locales que referencien los datashares. Especifique el NAMESPACE y el ID de la cuenta cuando cree la base de datos desde el recurso compartido de datos. Para obtener más información, consulte CREATE DATABASE.

```
CREATE DATABASE sales_db FROM DATASHARE salesshare OF ACCOUNT '123456789012'
 NAMESPACE 'dd8772e1-d792-4fa4-996b-1870577efc0d';
```

   Si desea tener un control más detallado del acceso a los objetos de la base de datos local, utilice la cláusula WITH PERMISSIONS al crear la base de datos. De este modo, podrá conceder permisos en el nivel de objeto para los objetos de la base de datos en el paso 4.

```
CREATE DATABASE sales_db WITH PERMISSIONS FROM DATASHARE salesshare OF ACCOUNT
 '123456789012' NAMESPACE 'dd8772e1-d792-4fa4-996b-1870577efc0d';
```

   Puede ver las bases de datos que creó a partir del datashare consultando la vista SVV_REDSHIFT_DATABASES. Puede conectarse a estas bases de datos directamente o puede conectarse a una base de datos local en el clúster del consumidor y realizar una consulta entre bases de datos para consultar los datos de las bases de datos de recursos compartidos de datos. No puede crear un datashare encima de los objetos de base de datos creados a partir de un datashare existente. No obstante, puede copiar los datos en una tabla diferente en el clúster consumidor, llevar a cabo el procesamiento que sea necesario y, a continuación, compartir los objetos nuevos creados.

3. (Opcional) Cree esquemas externos para referenciar esquemas específicos de la base de datos consumidora importados en el clúster consumidor y para asignarles permisos pormenorizados. Para obtener más información, consulte CREATE EXTERNAL SCHEMA.

```
CREATE EXTERNAL SCHEMA sales_schema FROM REDSHIFT DATABASE 'sales_db' SCHEMA
 'public';
```

4. Conceda permisos en las bases de datos y las referencias de esquema creadas a partir de los recursos compartidos de datos a usuarios o roles del clúster del consumidor según sea necesario. Para obtener más información, consulte GRANT o REVOKE.

```
GRANT USAGE ON DATABASE sales_db TO Bob;
```

```
GRANT USAGE ON SCHEMA sales_schema TO ROLE Analyst_role;
```

Si ha creado su base de datos sin WITH PERMISSIONS, solo podrá asignar permisos a toda la base de datos creada a partir del recurso compartido de datos a los usuarios o roles. En algunos casos, necesita controles pormenorizados sobre un subconjunto de los objetos de base de datos creados a partir del datashare. Si es así, puede crear una referencia de esquema externo que apunte a esquemas específicos del recurso compartido de datos, como se describe en el paso anterior. Luego, puede proporcionar permisos pormenorizados en el nivel de esquema. También puede crear vistas de enlace de tiempo de ejecución encima de los objetos compartidos y utilizarlas para asignar permisos pormenorizados. También puede considerar la posibilidad de indicar a los clústeres productores que creen datashares adicionales en su nombre con el nivel de detalle requerido. Puede crear tantas referencias de esquema para la base de datos creada a partir del recurso compartido de datos como quiera.

Si ha creado la base de datos con WITH PERMISSIONS en el paso 2, debe asignar permisos de nivel de objeto a los objetos de la base de datos compartida. Un usuario que solo tenga el permiso USAGE no podrá acceder a ningún objeto de una base de datos creada con WITH PERMISSIONS hasta que se le concedan permisos adicionales en el nivel de objeto.

```
GRANT SELECT ON sales_db.public.tickit_sales_redshift to Bob;
```

5. Consulte los datos de los objetos compartidos que se encuentran en los datashares.

   Los usuarios y los roles que tengan permisos en las bases de datos del consumidor y los esquemas de los clústeres del consumidor pueden explorar y navegar por los metadatos de cualquier objeto compartido. También pueden navegar por los objetos locales de un clúster consumidor y explorarlos. Para ello, utilice controladores JDBC u ODBC, o vistas SVV_ALL y SVV_REDSHIFT.

   Los clústeres productores pueden tener muchos esquemas en la base de datos y muchas tablas y vistas dentro de cada esquema. Los usuarios del lado del consumidor solo pueden ver el subconjunto de objetos que se han puesto a disposición a través del datashare. Estos usuarios no pueden ver todos los metadatos del clúster productor. Este enfoque ayuda a proporcionar un control pormenorizado de la seguridad de los metadatos con el uso compartido de datos.

   Continúa conectándose a las bases de datos locales del clúster. No obstante, ahora, también es posible leer de las bases de datos y los esquemas que se crean a partir de los datashares

utilizando la notación de tres partes database.schema.table. Puede realizar consultas que abarquen todas las bases de datos visibles para usted. Estas pueden ser bases de datos locales del clúster o bases de datos creadas a partir de los datashares. Los clústeres consumidores no pueden conectarse a las bases de datos que se hayan creado a partir de los datashares.

Puede acceder a los datos mediante la cualificación completa. Para obtener más información, consulte Ejemplos de consultas entre bases de datos.

```
SELECT * FROM sales_db.public.tickit_sales_redshift;
```

Solo se pueden utilizar las instrucciones SELECT en los objetos compartidos. No obstante, puede crear tablas en el clúster consumidor consultando los datos de los objetos compartidos de una base de datos local diferente.

Además de realizar consultas, los consumidores pueden crear vistas en los objetos compartidos. Solo se admiten las vistas de enlace en tiempo de ejecución o las vistas materializadas. Amazon Redshift no admite las vistas normales en los datos compartidos. Las vistas que crean los consumidores pueden abarcar varias bases de datos locales o bases de datos creadas a partir de datashares. Para obtener más información, consulte CREATE VIEW.

```
// Connect to a local cluster database

// Create a view on shared objects and access it.
CREATE VIEW sales_data
AS SELECT *
FROM sales_db.public.tickit_sales_redshift
WITH NO SCHEMA BINDING;

SELECT * FROM sales_data;
```

## Compartir datos entre Regiones de AWS

Puede compartir datos con fines de lectura entre los clústeres de Amazon Redshift en las Regiones de AWS. Con el uso compartido de datos entre regiones, puede compartir datos entre Regiones de AWS sin necesidad de copiar datos de forma manual. No tiene que descargar los datos en Amazon S3, copiarlos en un nuevo clúster de Amazon Redshift ni realizar copias de instantáneas entre regiones.

Con el uso compartido de datos entre regiones, puede compartir datos entre clústeres en la misma Cuenta de AWS o en diferentes Cuentas de AWS, incluso cuando los clústeres se encuentran en diferentes regiones. Cuando se comparten datos con clústeres de Amazon Redshift que se encuentran en la misma Cuenta de AWS, pero en diferentes Regiones de AWS, se debe seguir el mismo flujo de trabajo que para compartir datos dentro de Cuenta de AWS. Para obtener más información, consulte Uso compartido de acceso de lectura a los datos en una Cuenta de AWS.

Si los clústeres que comparten datos se encuentran en diferentes Cuentas de AWS y Regiones de AWS, se puede seguir el mismo flujo de trabajo que para compartir datos entre Cuentas de AWS e incluir asociaciones en el nivel de la región en el clúster consumidor. El uso compartido de datos entre regiones admite la asociación de recursos compartidos de datos con toda la Cuenta de AWS, toda la Región de AWS o espacios de nombres específicos dentro de una Región de AWS. Para obtener más información sobre el uso compartido de datos entre Cuentas de AWS, consulte Compartir datos entre Cuentas de AWS.

Cuando consume datos de una región diferente, el consumidor paga la tarifa de transferencia de datos entre regiones desde la región del productor a la región del consumidor.

Para utilizar el recurso compartido de datos, el administrador de una cuenta consumidora puede asociar el recurso compartido de datos de una de las siguientes tres formas:

- asociación con toda una Cuenta de AWS que abarca todas sus Regiones de AWS
- asociación con una determinada Región de AWS en una Cuenta de AWS
- Asociación con espacios de nombres específicos dentro de una Región de AWS

Cuando el administrador elige toda la Cuenta de AWS, todos los espacios de nombres existentes y futuros en diferentes Regiones de AWS de la cuenta tienen acceso a los recursos compartidos de datos. El administrador de una cuenta de consumidor también puede elegir Regiones de AWS o espacios de nombres específicos dentro de una región para concederles acceso a los recursos compartidos de datos.

Si es administrador de productores o el propietario de la base de datos, cree un recurso compartido de datos, añada objetos de base de datos y consumidores de datos al recurso compartido de datos, y conceda permisos a los consumidores de datos. Para obtener más información, consulte Acciones del administrador de productores.

Si es administrador de una cuenta productora, autorice los recursos compartidos de datos mediante la AWS Command Line Interface (AWS CLI) o la consola de Amazon Redshift y elija los consumidores de datos.

Si es el administrador de una cuenta consumidora, siga estos pasos:

Para asociar uno o más recursos compartidos de datos que se comparten desde otras cuentas con toda la Cuenta de AWS o con Regiones de AWS o espacios de nombres específicos dentro de una Región de AWS, utilice la consola de Amazon Redshift.

Con el uso compartido de datos entre regiones, puede añadir clústeres en una Región de AWS específica mediante la AWS Command Line Interface (AWS CLI) o la consola de Amazon Redshift.

Para especificar una o varias regiones de AWS, puede utilizar el comando de la CLI `associate-data-share-consumer` con el ajuste `consumer-region` opcional.

Con la CLI, en el siguiente ejemplo se asocia `Salesshare` a toda la Cuenta de AWS con la opción `associate-entire-account`. Solo puede asociar una región a la vez.

```
aws redshift associate-data-share-consumer
--region {PRODUCER_REGION}
--data-share-arn arn:aws:redshift:{PRODUCER_REGION}:{PRODUCER_ACCOUNT}:datashare:
{PRODUCER_CLUSTER_NAMESPACE}/Salesshare
--associate-entire-account
```

El siguiente ejemplo asocia `Salesshare` con la región Este de EE. UU. (Ohio) (`us-east-2`).

```
aws redshift associate-data-share-consumer
--region {PRODUCER_REGION}
--data-share-arn arn:aws:redshift:{PRODUCER_REGION}:0123456789012:datashare:
{PRODUCER_CLUSTER_NAMESPACE}/Salesshare
--consumer-region 'us-east-2'
```

En el siguiente ejemplo, se asocia `Salesshare` con un espacio de nombres de consumidor específico de otra Cuenta de AWS de la región Asia-Pacífico (Sídney) (`ap-southeast-2`).

```
aws redshift associate-data-share-consumer
--data-share-arn arn:aws:redshift:{PRODUCER_REGION}:{PRODUCER_ACCOUNT}:datashare:
{PRODUCER_CLUSTER_NAMESPACE}/Salesshare
--consumer-arn 'arn:aws:redshift:ap-southeast-2:{CONSUMER_ACCOUNT}:namespace:
{ConsumerImmutableClusterId}'
```

Puede utilizar la consola de Amazon Redshift para asociar recursos compartidos de datos con toda la Cuenta de AWS o con Regiones de AWS o espacios de nombres específicos dentro de una Región de AWS. Para ello, inicie sesión en https://console.aws.amazon.com/redshiftv2/. Luego, asocie uno o más recursos compartidos de datos que se comparten desde otras cuentas con toda la Cuenta de AWS, con toda la Región de AWS o con un espacio de nombres específico en una Región de AWS. Para obtener más información, consulte Asociación de un recurso compartido de datos desde otra Cuenta de AWS en Amazon Redshift.

Después de asociar la Cuenta de AWS o espacios de nombres específicos, los recursos compartidos de datos están a su disposición para que los consuma. También puede cambiar la asociación del datashare en cualquier momento. Cuando se cambia la asociación de espacios de nombres individuales a una Cuenta de AWS, Amazon Redshift sobrescribe los espacios de nombres con la información de la Cuenta de AWS. Cuando se cambia la asociación de una Cuenta de AWS a espacios de nombres específicos, Amazon Redshift sobrescribe la información de la Cuenta de AWS con la información del espacio de nombres. Cuando se cambia la asociación de toda una Cuenta de AWS a regiones de AWS y espacios de nombres específicos, Amazon Redshift sobrescribe la información de la Cuenta de AWS con la información de la región y el espacio de nombres específicos.

Si es administrador de consumidores, puede crear bases de datos locales que hagan referencia a los recursos compartidos de datos y conceder permisos en las bases de datos creadas a partir de los recursos compartidos de datos a los usuarios o roles en el clúster del consumidor según sea necesario. También puede crear vistas en los objetos compartidos y esquemas externos para referenciar esquemas específicos de la base de datos consumidora importados en el clúster consumidor y para asignarles permisos pormenorizados. Para obtener más información, consulte Acciones del administrador de consumidores.

Administración del control de costos para el uso compartido de datos entre regiones

Con Amazon Redshift, puede administrar el control de costos para el uso compartido de datos entre regiones al configurar el uso compartido de datos para limitar la cantidad de datos que se transfieren entre regiones de AWS. La administración del control de los costos del uso compartido de datos entre regiones le permite establecer límites de transferencia de datos, supervisar el uso de la transferencia de datos y recibir notificaciones cuando se acerque o supere esos límites.

Al consumir datos de una región diferente, el consumidor paga la tarifa de transferencia de datos entre regiones de la región productora a la región de consumo. El precio de la transferencia de datos es diferente para las distintas regiones. El cargo se basa en los bytes de datos escaneados para

cada consulta que se ejecuta con éxito. Para obtener más información sobre los precios de Amazon Redshift, consulte Precios de Amazon Redshift.

Se le cobra el número de bytes, redondeado al siguiente megabyte, con un mínimo de 10 MB por consulta. Puede establecer controles de costo sobre el uso de las consultas y ver la cantidad de datos que se transfieren por cada consulta en el clúster.

Para supervisar y controlar el uso y el costo asociado al uso compartido de datos entre regiones, puede crear límites de uso diarios, semanales y mensuales, y definir acciones que Amazon Redshift deba realizar automáticamente si se alcanzan esos límites, para ayudar a mantener un presupuesto previsible.

En función de los límites de uso que se establezcan, las acciones que Amazon Redshift realiza pueden ser registrar un evento en una tabla del sistema, enviar una alarma de CloudWatch y notificar a un administrador con Amazon SNS, o desactivar el uso compartido de datos entre regiones para un uso posterior.

Para crear límites de uso en la consola de Amazon Redshift, elija Configure usage limit (Configurar límite de uso) en Actions (Acciones) para el clúster. Puede supervisar las tendencias de uso y recibir alertas cuando el uso supere los límites definidos con métricas de CloudWatch generadas automáticamente desde las pestañas Cluster performance (Rendimiento del clúster) o Monitoring (Supervisión). También puede crear, modificar y eliminar límites de uso mediante programación utilizando la AWS CLI u operaciones de la API de Amazon Redshift.

## Uso compartido de datos de Amazon Redshift con licencia en AWS Data Exchange

Cuando se crean recursos compartidos de datos de AWS Data Exchange y se agregan a un producto de AWS Data Exchange, los proveedores pueden utilizar licencias de datos en Amazon Redshift. Los consumidores pueden descubrir y consultar estos datos actualizados, además de suscribirse a ellos, en Amazon Redshift cuando tienen suscripciones activas de AWS Data Exchange.

Con los recursos compartidos de datos de AWS Data Exchange agregados a un producto de AWS Data Exchange, los consumidores tienen acceso de forma automática a los recursos compartidos de datos de un producto cuando comienza su suscripción y conservan su acceso mientras su suscripción esté activa.

Temas

- Uso de recursos compartidos de datos de AWS Data Exchange como productor
- Uso de recursos compartidos de datos de AWS Data Exchange como consumidor

### Uso de recursos compartidos de datos de AWS Data Exchange como productor

Con Amazon Redshift, puede compartir productos de datos en directo con AWS Data Exchange como un productor que crea y administra recursos compartidos de datos.

Si es administrador de productores, siga estos pasos para administrar los recursos compartidos de datos de AWS Data Exchange en la consola de Amazon Redshift:

1. Cree recursos compartidos de datos en el clúster para compartir datos en AWS Data Exchange y otorgue acceso a AWS Data Exchange a estos recursos.

   Los superusuarios de clústeres y los propietarios de bases de datos pueden crear datashares. Cada datashare se asocia a una base de datos durante la creación. Solo los objetos de esa base de datos se pueden compartir en ese datashare. Se pueden crear varios datashares en la misma base de datos con el mismo nivel de detalle de los objetos o uno diferente. No hay límite en cuanto a la cantidad de recursos compartidos de datos que puede crear en un clúster.

   También puede utilizar la consola de Amazon Redshift para crear datashares. Para obtener más información, consulte Creación de un recurso compartido de datos.

   Utilice la opción MANAGEDBY ADX para otorgar acceso de forma implícita al recurso compartido de datos a AWS Data Exchange cuando se ejecute la instrucción CREATE DATASHARE. Esto indica que AWS Data Exchange administra este recurso compartido de datos. Solo puede utilizar la opción MANAGEDBY ADX cuando crea un recurso compartido de datos nuevo. No puede utilizar la instrucción ALTER DATASHARE para modificar un recurso compartido de datos existente para agregar la opción MANAGEDBY ADX. Una vez que se crea el recurso compartido de datos con la opción MANAGEDBY ADX, solo AWS Data Exchange puede acceder a este recurso y administrarlo.

   ```
   CREATE DATASHARE salesshare
   [[SET] MANAGEDBY [=] {ADX} ];
   ```

2. Agregue objetos a los recursos compartidos de datos. El administrador del productor continúa administrando los objetos del recurso compartido de datos que están disponibles en un recurso compartido de datos de AWS Data Exchange.

   Para agregar objetos a un datashare, agregue el esquema antes de agregar los objetos. Cuando se agrega un esquema, Amazon Redshift no agrega todos los objetos que contiene. Se deben agregar de forma explícita. Para obtener más información, consulte ALTER DATASHARE.

```
ALTER DATASHARE salesshare ADD SCHEMA PUBLIC;
ALTER DATASHARE salesshare ADD TABLE public.tickit_sales_redshift;
ALTER DATASHARE salesshare ADD ALL TABLES IN SCHEMA PUBLIC;
```

También se pueden agregar vistas a un datashare.

```
CREATE VIEW public.sales_data_summary_view AS SELECT * FROM
 public.tickit_sales_redshift;
ALTER DATASHARE salesshare ADD TABLE public.sales_data_summary_view;
```

Utilice ALTER DATASHARE para compartir esquemas, tablas, vistas y funciones de un esquema determinado. Los superusuarios, los propietarios de recursos compartidos de datos o los usuarios que tienen permisos ALTER u ALL sobre el recurso compartido de datos pueden modificarlo para agregarle o quitarle objetos. Los usuarios deben tener los permisos para poder agregar objetos al recurso compartido de datos o quitarlos. Los usuarios también deben ser los propietarios de los objetos o disponer de permisos SELECT, USAGE u ALL sobre ellos.

Utilice la cláusula INCLUDENEW para agregar al datashare nuevas tablas, vistas o funciones definidas por el usuario (UDF) de SQL que se crearon en un esquema especificado. Solo los superusuarios pueden modificar esta propiedad para cada par de esquema y datashare.

```
ALTER DATASHARE salesshare ADD SCHEMA PUBLIC;
ALTER DATASHARE salesshare SET INCLUDENEW = TRUE FOR SCHEMA PUBLIC;
```

También puede utilizar la consola de Amazon Redshift para agregar objetos a los datashares o quitarlos. Para obtener más información, consulte Adición de objetos de recursos compartidos de datos a recursos compartidos de datos, Eliminación de objetos de los datashares y Edición de los conjuntos compartidos de datos en AWS Data Exchange.

3. Para autorizar el acceso a los recursos compartidos de datos de AWS Data Exchange, realice una de las siguientes acciones:

   - Autorice el acceso al recurso compartido de datos para AWS Data Exchange de forma explícita utilizando la palabra clave ADX en la API `aws redshift authorize-data-share`. Esto permite que AWS Data Exchange reconozca el recurso compartido de datos en la cuenta de servicio y administre la asociación de los consumidores al recurso compartido de datos.

```
aws redshift authorize-data-share
--data-share-arn arn:aws:redshift:us-east-1:{PRODUCER_ACCOUNT}:datashare:
{PRODUCER_CLUSTER_NAMESPACE}/salesshare
--consumer-identifier ADX
```

Puede utilizar la clave condicional `ConsumerIdentifier` para las API `AuthorizeDataShare` y `DeauthorizeDataShare` para permitir o denegar de forma explícita que AWS Data Exchange realice llamadas a las dos API en la política de IAM.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "VisualEditor0",
            "Effect": "Deny",
            "Action": [
                "redshift:AuthorizeDataShare",
                "redshift:DeauthorizeDataShare"
            ],
            "Resource": "*",
            "Condition": {
                "StringEqualsIgnoreCase": {
                    "redshift:ConsumerIdentifier": "ADX"
                }
            }
        }
    ]
}
```

- Utilice la consola de Amazon Redshift para otorgar o eliminar la autorización de los recursos compartidos de datos de AWS Data Exchange. Para obtener más información, consulte Autorización o retiro de la autorización de los recursos compartidos de datos.
- Opcionalmente, puede autorizar de forma implícita el acceso al recurso compartido de datos de AWS Data Exchange cuando importe el recurso compartido de datos a un conjunto de datos de AWS Data Exchange.

Para eliminar la autorización de acceso a los recursos compartidos de datos de AWS Data Exchange, utilice la palabra clave `ADX` en la operación de la API `aws redshift deauthorize-data-share`. Esto permite que AWS Data Exchange reconozca el recurso

compartido de datos en la cuenta de servicio y administre la eliminación de la asociación del recurso compartido de datos.

```
aws redshift deauthorize-data-share
--data-share-arn arn:aws:redshift:us-east-1:{PRODUCER_ACCOUNT}:datashare:
{PRODUCER_CLUSTER_NAMESPACE}/salesshare
--consumer-identifier ADX
```

4. Enumere los datashares que se hayan creado en el clúster y examine su contenido.

   En el siguiente ejemplo, se muestra la información de un recurso compartido de datos denominado SalesShare. Para obtener más información, consulte DESC DATASHARE y SHOW DATASHARES.

```
DESC DATASHARE salesshare;

 producer_account  |          producer_namespace          | share_type | share_name
 | object_type |           object_name          |   include_new
------------------+--------------------------------------+------------
+------------+-------------+---------------------------------+------------------
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | table       | public.tickit_users_redshift   |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | table       | public.tickit_venue_redshift   |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | table       | public.tickit_category_redshift|
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | table       | public.tickit_date_redshift    |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | table       | public.tickit_event_redshift   |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | table       | public.tickit_listing_redshift |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | table       | public.tickit_sales_redshift   |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | schema      | public                         |  t
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | view        | public.sales_data_summary_view |
```

   En el siguiente ejemplo, se muestran los datashares salientes de un clúster productor.

```
SHOW DATASHARES LIKE 'sales%';
```

El resultado tiene un aspecto similar al siguiente.

```
share_name | share_owner  | source_database | consumer_database | share_type |
 createdate      | is_publicaccessible  | share_acl | producer_account |
 producer_namespace
----------+-------------+-----------------+-------------------+------------
+---------------------+----------------------+-----------+-----------------
+--------------------------------------
salesshare |    100       | dev             |                   |  OUTBOUND
 | 2020-12-09 02:27:08 |          True        |           |   123456789012   |
 13b8833d-17c6-4f16-8fe4-1a018f5ed00d
```

Para obtener más información, consulte DESC DATASHARE y SHOW DATASHARES.

También puede utilizar SVV_DATASHARES, SVV_DATASHARE_CONSUMERS y SVV_DATASHARE_OBJECTS para ver los datashares, los objetos dentro del datashare y sus consumidores.

5. Elimine datashares. Le recomendamos que no elimine un recurso compartido de datos de AWS Data Exchange compartido con otras Cuentas de AWS mediante la instrucción DROP DATASHARE. Esas cuentas perderán el acceso al recurso compartido de datos. Esta acción es irreversible. Esto podría infringir las condiciones de la oferta del producto de datos en AWS Data Exchange. Si desea borrar un recurso compartido de datos de AWS Data Exchange, consulte Notas de uso de DROP DATASHARE.

   En el siguiente ejemplo, se elimina un recurso compartido de datos denominado SalesShare.

   ```
   DROP DATASHARE salesshare;
   ERROR:  Drop of ADX-managed datashare salesshare requires session variable
    datashare_break_glass_session_var to be set to value '620c871f890c49'
   ```

   Para permitir que se elimine un recurso compartido de datos de AWS Data Exchange, establezca la variable datashare_break_glass_session_var y ejecute de nuevo la instrucción DROP DATASHARE. Si desea borrar un recurso compartido de datos de AWS Data Exchange, consulte Notas de uso de DROP DATASHARE.

También puede utilizar la consola de Amazon Redshift para eliminar los datashares. Para obtener más información, consulte Eliminación de los recursos compartidos de datos de AWS Data Exchange creados en su cuenta.

6. Utilice ALTER DATASHARE para quitar objetos de los datashares en cualquier parte de este. Utilice REVOKE USAGE ON para revocar los permisos sobre el datashare otorgados a determinados consumidores. Revoca los permisos USAGE de los objetos que se encuentran dentro de un recurso compartido de datos y desactiva de forma instantánea el acceso para todos los clústeres consumidores. Las listas de los datashares y las consultas de metadatos, como las listas de bases de datos y tablas, no devuelven los objetos compartidos después de haber revocado el acceso.

```
ALTER DATASHARE salesshare REMOVE TABLE public.tickit_sales_redshift;
```

También puede utilizar la consola de Amazon Redshift para modificar datashares. Para obtener más información, consulte Edición de los conjuntos compartidos de datos en AWS Data Exchange.

7. Conceda o revoque GRANT USAGE de los recursos compartidos de datos de AWS Data Exchange. No puede conceder ni revocar GRANT USAGE para el recurso compartido de datos de AWS Data Exchange. En el siguiente ejemplo, se muestra un error cuando se concede el permiso GRANT USAGE a una Cuenta de AWS para un recurso compartido de datos que administra AWS Data Exchange.

```
CREATE DATASHARE salesshare MANAGEDBY ADX;
```

```
GRANT USAGE ON DATASHARE salesshare TO ACCOUNT '012345678910';
ERROR:  Permission denied to add/remove consumer to/from datashare salesshare.
 Datashare consumers are managed by ADX.
```

Para obtener más información, consulte GRANT o REVOKE.

Si es administrador de productores, siga estos pasos para crear y publicar un producto de recursos compartidos de datos en la consola de AWS Data Exchange:

- Una vez que se crea el recurso compartido de datos de AWS Data Exchange, el productor crea un conjunto de datos nuevo, importa activos, crea una revisión, y crea y publica un producto nuevo.

  Utilice la consola de Amazon Redshift para crear conjuntos de datos. Para obtener más información, consulte Creación de conjuntos de datos en AWS Data Exchange.

  Para obtener más información, consulte Proporcionar productos de datos en AWS Data Exchange.

Uso de recursos compartidos de datos de AWS Data Exchange como consumidor

Con Amazon Redshift, puede acceder a los conjuntos de datos y analizarlos desde AWS Data Exchange sin tener que almacenar ni administrar copias de los datos.

Si es consumidor, siga estos pasos para descubrir productos de datos que contengan recurso compartido de datos de AWS Data Exchange y consultar los datos de Amazon Redshift:

1. En la consola de AWS Data Exchange, descubra productos de datos que contengan recursos compartidos de datos de AWS Data Exchange y suscríbase a ellos.

   Una vez que se inicie la suscripción, podrá acceder a los datos de Amazon Redshift con licencia que se importan como activos a conjuntos de datos que contienen recursos compartidos de datos de AWS Data Exchange.

   Para obtener más información sobre cómo empezar a utilizar productos de datos que contengan recursos compartidos de datos de AWS Data Exchange, consulte Suscripción a productos de datos en AWS Data Exchange.
2. En la consola de Amazon Redshift, cree un clúster de Amazon Redshift, en caso de ser necesario.

   Para obtener información sobre cómo crear un clúster, consulte Creación de un clúster.
3. Enumere los datashares que se ponen a su disposición y vea su contenido. Para obtener más información, consulte DESC DATASHARE y SHOW DATASHARES.

   En el siguiente ejemplo, se muestra la información de los datashares entrantes de un espacio de nombres productor específico. Cuando ejecuta DESC DATASHARE como administrador de consumidores, debe especificar la opción ACCOUNT y NAMESPACE para ver los recursos compartidos de datos entrantes.

```
DESC DATASHARE salesshare of ACCOUNT '123456789012' NAMESPACE
 '13b8833d-17c6-4f16-8fe4-1a018f5ed00d';

 producer_account  |          producer_namespace          | share_type | share_name
 | object_type |           object_name           |   include_new
-------------------+--------------------------------------+------------
+------------+-------------+----------------------------------+-----------------
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    | salesshare
 | table       | public.tickit_users_redshift    |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    | salesshare
 | table       | public.tickit_venue_redshift    |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    | salesshare
 | table       | public.tickit_category_redshift |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    | salesshare
 | table       | public.tickit_date_redshift     |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    | salesshare
 | table       | public.tickit_event_redshift    |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    | salesshare
 | table       | public.tickit_listing_redshift  |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    | salesshare
 | table       | public.tickit_sales_redshift    |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    | salesshare
 | schema      | public                          |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    | salesshare
 | view        | public.sales_data_summary_view  |
```

Solo los superusuarios del clúster pueden hacer esto. También puede utilizar SVV_DATASHARES para ver los datashares y SVV_DATASHARE_OBJECTS para ver los objetos que se encuentran dentro del datashare.

En el siguiente ejemplo, se muestran los datashares entrantes de un clúster consumidor.

```
SHOW DATASHARES LIKE 'sales%';


 share_name | share_owner | source_database | consumer_database | share_type
 | createdate | is_publicaccessible | share_acl | producer_account |
 producer_namespace
------------+-------------+-----------------+-------------------+------------
+------------+---------------------+-----------+------------------
+--------------------------------------
```

```
 salesshare |             |                 |                   | INBOUND
    |            |         t          |           |   123456789012   |
 13b8833d-17c6-4f16-8fe4-1a018f5ed00d
```

4. Cree bases de datos locales que referencien los datashares. Debe especificar la opción ACCOUNT y NAMESPACE para crear bases de datos locales para los recursos compartidos de datos de AWS Data Exchange. Para obtener más información, consulte CREATE DATABASE.

```
CREATE DATABASE sales_db FROM DATASHARE salesshare OF ACCOUNT '123456789012'
 NAMESPACE '13b8833d-17c6-4f16-8fe4-1a018f5ed00d';
```

   Si desea tener un control más detallado del acceso a los objetos de la base de datos local, utilice la cláusula WITH PERMISSIONS al crear la base de datos. De este modo, podrá conceder permisos en el nivel de objeto para los objetos de la base de datos en el paso 6.

```
CREATE DATABASE sales_db WITH PERMISSIONS FROM DATASHARE salesshare OF ACCOUNT
 '123456789012' NAMESPACE '13b8833d-17c6-4f16-8fe4-1a018f5ed00d';
```

   Puede ver las bases de datos que creó a partir del datashare consultando la vista SVV_REDSHIFT_DATABASES. Puede conectarse a estas bases de datos directamente o puede conectarse a una base de datos local en el clúster del consumidor y realizar una consulta entre bases de datos para consultar los datos de las bases de datos de recursos compartidos de datos. No puede crear un datashare encima de los objetos de base de datos creados a partir de un datashare existente. No obstante, puede copiar los datos en una tabla diferente en el clúster consumidor, llevar a cabo el procesamiento que sea necesario y, a continuación, compartir los objetos nuevos que se hayan creado.

   También puede utilizar la consola de Amazon Redshift para crear bases de datos a partir de datashares. Para obtener más información, consulte Creación de bases de datos a partir de datashares.

5. (Opcional) Cree esquemas externos para referenciar esquemas específicos de la base de datos consumidora importados en el clúster consumidor y para asignarles permisos pormenorizados. Para obtener más información, consulte CREATE EXTERNAL SCHEMA.

```
CREATE EXTERNAL SCHEMA sales_schema FROM REDSHIFT DATABASE 'sales_db' SCHEMA
 'public';
```

6. Conceda permisos en las bases de datos y las referencias de esquema creadas a partir de los recursos compartidos de datos a usuarios o roles del clúster del consumidor según sea necesario. Para obtener más información, consulte GRANT o REVOKE.

```
GRANT USAGE ON DATABASE sales_db TO Bob;
```

```
GRANT USAGE ON SCHEMA sales_schema TO ROLE Analyst_role;
```

Si ha creado su base de datos sin WITH PERMISSIONS, solo podrá asignar permisos a toda la base de datos creada a partir del recurso compartido de datos a los usuarios y los roles. En algunos casos, necesita controles pormenorizados sobre un subconjunto de los objetos de base de datos creados a partir del datashare. Si es así, puede crear una referencia de esquema externa que apunte a esquemas específicos del datashare (como se describe en el paso anterior) y proporcionar permisos pormenorizados en el nivel del esquema.

También puede crear vistas de enlace de tiempo de ejecución encima de los objetos compartidos y utilizarlas para asignar permisos pormenorizados. También puede considerar la posibilidad de indicar a los clústeres productores que creen datashares adicionales en su nombre con el nivel de detalle requerido. Puede crear tantas referencias de esquema para la base de datos creada a partir del recurso compartido de datos como quiera.

Si ha creado la base de datos con WITH PERMISSIONS en el paso 4, debe asignar permisos de nivel de objeto a los objetos de la base de datos compartida. Un usuario que solo tenga el permiso USAGE no podrá acceder a ningún objeto de una base de datos creada con WITH PERMISSIONS hasta que se le concedan permisos adicionales en el nivel de objeto.

```
GRANT SELECT ON sales_db.public.tickit_sales_redshift to Bob;
```

7. Consulte los datos de los objetos compartidos que se encuentran en los datashares.

Los usuarios y los roles que tengan permisos en las bases de datos del consumidor y los esquemas de los clústeres del consumidor pueden explorar y navegar por los metadatos de cualquier objeto compartido. También pueden navegar por los objetos locales de un clúster consumidor y explorarlos. Para ello, utilizan controladores JDBC u ODBC, los comandos SHOW o vistas SVV_ALL y SVV_REDSHIFT.

Los clústeres productores pueden tener muchos esquemas en la base de datos y muchas tablas y vistas dentro de cada esquema. Los usuarios del lado del consumidor solo pueden ver el

subconjunto de objetos que se han puesto a disposición a través del datashare. Estos usuarios no pueden ver todos los metadatos del clúster productor. Este enfoque ayuda a proporcionar un control pormenorizado de la seguridad de los metadatos con el uso compartido de datos.

Continúa conectándose a las bases de datos locales del clúster. No obstante, ahora, también es posible leer de las bases de datos y los esquemas que se crean a partir de los datashares utilizando la notación de tres partes database.schema.table. Puede realizar consultas que abarquen todas las bases de datos visibles para usted. Estas pueden ser bases de datos locales del clúster o bases de datos creadas a partir de los datashares. O bien, puede conectarse directamente a estas bases de datos de consumidores y ejecutar consultas en los objetos compartidos con una notación parcial.

Puede acceder a los datos mediante la cualificación completa. Para obtener más información, consulte Ejemplos de consultas entre bases de datos.

```
SELECT * FROM sales_db.public.tickit_sales_redshift ORDER BY 1,2 LIMIT 5;

 salesid | listid | sellerid | buyerid | eventid | dateid | qtysold | pricepaid |
 commission |      saletime
---------+--------+----------+---------+---------+--------+---------+-----------
+------------+---------------------
       1 |      1 |    36861 |   21191 |    7872 |   1875 |       4 |    728.00 |
   109.20 | 2008-02-18 02:36:48
       2 |      4 |     8117 |   11498 |    4337 |   1983 |       2 |     76.00 |
    11.40 | 2008-06-06 05:00:16
       3 |      5 |     1616 |   17433 |    8647 |   1983 |       2 |    350.00 |
    52.50 | 2008-06-06 08:26:17
       4 |      5 |     1616 |   19715 |    8647 |   1986 |       1 |    175.00 |
    26.25 | 2008-06-09 08:38:52
       5 |      6 |    47402 |   14115 |    8240 |   2069 |       2 |    154.00 |
    23.10 | 2008-08-31 09:17:02
```

Solo se pueden utilizar las instrucciones SELECT en los objetos compartidos. No obstante, puede crear tablas en el clúster consumidor consultando los datos de los objetos compartidos de una base de datos local diferente.

Además de las consultas, los consumidores pueden crear vistas en los objetos compartidos. Solo se admiten las vistas de enlace de tiempo de ejecución o las vistas materializadas. Amazon Redshift no admite las vistas normales en los datos compartidos. Las vistas que crean

los consumidores pueden abarcar varias bases de datos locales o bases de datos creadas a partir de datashares. Para obtener más información, consulte CREATE VIEW.

```
// Connect to a local cluster database

// Create a view on shared objects and access it.
CREATE VIEW sales_data
AS SELECT *
FROM sales_db.public.tickit_sales_redshift
WITH NO SCHEMA BINDING;

SELECT * FROM sales_data;
```

## Introducción a los recursos compartidos de datos administrados por AWS Lake Formation

Con Amazon Redshift, puede acceder a los datos en directo entre cuentas de AWS y clústeres de Amazon Redshift y compartirlos mediante recursos compartidos de datos administrados por AWS Lake Formation. Los recursos compartidos de datos de AWS Lake Formation permiten a los proveedores de datos compartir de forma segura los datos en directo del lago de datos de Amazon S3 con cualquier consumidor, incluidas otras cuentas de AWS y clústeres de Amazon Redshift.

Uso de recursos compartidos de datos administrados por Lake Formation como productor

Con Amazon Redshift, puede acceder a los datos compartidos a través de recursos compartidos de datos de AWS Lake Formation y analizarlos. Los recursos compartidos de datos de AWS Lake Formation permiten proteger el uso compartido de datos en las cuentas de AWS y clústeres de Amazon Redshift sin tener que copiar ni mover los datos subyacentes.

Compartir datos con AWS Lake Formation le permite definir de forma centralizada los permisos de AWS Lake Formation de los recursos compartidos de datos de Amazon Redshift y restringir el acceso de los usuarios a los objetos de un recurso compartido de datos.

Con Amazon Redshift, puede compartir datos en directo de forma segura entre las cuentas de AWS y los clústeres de Amazon Redshift con recursos compartidos de datos administrados por AWS Lake Formation como productor. Un recurso compartido de datos administrado por Lake Formation es un objeto que le permite compartir datos en directo del clúster de Amazon Redshift con otras cuentas y servicios de AWS.

Como administrador del clúster o grupo de trabajo del productor, siga estos pasos para compartir los recursos compartidos de datos con Lake Formation:

1. Cree recursos compartidos de datos en el clúster y autorice el acceso a los recursos compartidos de datos a AWS Lake Formation.

   Los superusuarios de clústeres y los propietarios de bases de datos pueden crear recursos compartidos de datos. Cada datashare se asocia a una base de datos durante la creación. Solo los objetos de esa base de datos se pueden compartir en ese datashare. Se pueden crear varios datashares en la misma base de datos con el mismo nivel de detalle de los objetos o uno diferente. No hay límite en cuanto a la cantidad de recursos compartidos de datos que puede crear en un clúster.

   ```
   CREATE DATASHARE salesshare;
   ```

2. Agregue objetos al recurso compartido de datos. El administrador del clúster o grupo de trabajo del productor sigue administrando los objetos del recurso compartido de datos que estén disponibles. Para agregar objetos a un datashare, agregue el esquema antes de agregar los objetos. Cuando se agrega un esquema, Amazon Redshift no agrega todos los objetos que contiene. Se deben agregar de forma explícita. Para obtener más información, consulte ALTER DATASHARE.

   ```
   ALTER DATASHARE salesshare ADD SCHEMA PUBLIC;
   ALTER DATASHARE salesshare ADD TABLE public.tickit_sales_redshift;
   ALTER DATASHARE salesshare ADD ALL TABLES IN SCHEMA PUBLIC;
   ```

   También se pueden agregar vistas a un datashare. Los tipos de vistas que se admiten son: estándar, enlace en tiempo de ejecución y materializadas.

   ```
   CREATE VIEW public.sales_data_summary_view AS SELECT * FROM
    public.tickit_sales_redshift;
   ALTER DATASHARE salesshare ADD TABLE public.tickit_sales_redshift;
   ```

   Utilice ALTER DATASHARE para compartir esquemas, tablas y vistas de un esquema determinado. Los superusuarios, los propietarios de recursos compartidos de datos o los usuarios que tienen permisos ALTER u ALL sobre el recurso compartido de datos pueden modificarlo para agregarle o quitarle objetos. Los usuarios de base de datos también deben ser los propietarios de los objetos o disponer de permisos SELECT, USAGE u ALL sobre ellos.

Utilice la cláusula INCLUDENEW para agregar al recurso compartido de datos nuevas tablas y vistas creadas en un esquema especificado. Solo los superusuarios pueden modificar esta propiedad para cada par de esquema y datashare.

```
ALTER DATASHARE salesshare ADD SCHEMA PUBLIC;
ALTER DATASHARE salesshare SET INCLUDENEW = TRUE FOR SCHEMA PUBLIC;
```

3. Conceda acceso al recurso compartido de datos a una cuenta de administrador de Lake Formation.

```
GRANT USAGE ON DATASHARE salesshare TO ACCOUNT '012345678910' VIA DATA CATALOG;
```

Para revocar el uso, ejecute el siguiente comando.

```
REVOKE USAGE ON DATASHARE salesshare FROM ACCOUNT '012345678910' VIA DATA CATALOG;
```

4. Autorice el acceso al recurso compartido de datos de Lake Formation mediante la operación de la API `aws redshift authorize-data-share`. Al hacerlo, Lake Formation reconoce el recurso compartido de datos en la cuenta de servicio y administra la asociación de los consumidores al recurso compartido de datos.

```
aws redshift authorize-data-share
--data-share-arn arn:aws:redshift:us-east-1:{PRODUCER_ACCOUNT}:datashare:
{PRODUCER_CLUSTER_NAMESPACE}/salesshare
--consumer-identifier {"DataCatalog/<consumer-account-id>"}
```

Para eliminar la autorización de los recursos compartidos de datos administrado por Lake Formation, utilice la operación de la API `aws redshift deauthorize-data-share`. Al hacerlo así, permite a AWS Lake Formation que reconozca el recurso compartido de datos en la cuenta de servicio y retire la autorización.

```
aws redshift deauthorize-data-share
--data-share-arn arn:aws:redshift:us-east-1:{PRODUCER_ACCOUNT}:datashare:
{PRODUCER_CLUSTER_NAMESPACE}/salesshare
--consumer-identifier {"DataCatalog/<consumer-account-id>"}
```

En cualquier momento, si el administrador del clúster o grupo de trabajo del productor decide que ya no es necesario compartir datos con el clúster o el grupo de trabajo del consumidor,

puede usar DROP DATASHARE para eliminar el recurso compartido de datos, desautorizar el recurso compartido de datos o revocar los permisos del recurso compartido de datos. Los permisos y los objetos asociados en Lake Formation no se eliminan automáticamente.

```
DROP DATASHARE salesshare;
```

Tras autorizar a la cuenta de Lake Formation para administrar el recurso compartido de datos, el administrador de Lake Formation puede descubrir el recurso compartido de datos que se ha compartido, asociar el recurso compartido de datos con un ARN de catálogo de datos y crear una base de datos en AWS Glue Data Catalog que enlaza con el recurso compartido de datos. Para asociar recursos compartidos de datos mediante la AWS CLI, use el comando associate-data-share-consumer. Para compartir un recurso compartido de datos entre Regiones de AWS, especifique el parámetro `--region` en el comando `associate-data-share-consumer` o utilice la consola de AWS para elegir a los consumidores de datos. En el siguiente ejemplo se muestra cómo compartir un recurso compartido de datos administrado por Lake Formation entre regiones.

```
aws redshift associate-data-share-consumer --region <region-1>
--data-share-arn 'arn:aws:redshift:us-
east-1:12345678912:datashare:035c45ea-61ce-86f0-8b75-19ac6102c3b7/sample_share'
--consumer-arn 'arn:aws:glue:<region-1>:111912345678:catalog'
```

El administrador de Lake Formation también debe crear recursos locales que definan cómo los objetos del recurso compartido de datos deben asignarse a los objetos de Lake Formation. Para obtener más información sobre el descubrimiento de recursos compartidos de datos y la creación de recursos locales, consulte Administración de permisos para datos en un recurso compartido de datos de Amazon Redshift.

### Uso de recursos compartidos de datos administrados por Lake Formation como consumidor

Con Amazon Redshift, puede acceder y analizar datos compartidos con usted a través de recursos compartidos de datos de AWS Lake Formation. Un recurso compartido de datos es un producto de datos que contiene una recopilación de objetos de datos, como tablas o bases de datos, de diferentes orígenes de datos.

Una vez que el administrador de AWS Lake Formation descubra la invitación al recurso compartido de datos y cree una base de datos en el AWS Glue Data Catalog que se vincule con el recurso compartido de datos, el administrador del clúster o grupo de trabajo del consumidor puede asociar

el clúster al recurso compartido de datos y a la base de datos del AWS Glue Data Catalog, crear una base de datos local para el clúster o grupo de trabajo del consumidor y conceder acceso a los usuarios y los roles del clúster o grupo de trabajo del consumidor de Amazon Redshift para comenzar a hacer consultas. Siga estos pasos para configurar los permisos de consulta.

1. En la consola de Amazon Redshift, cree un clúster de Redshift que actúe como clúster o grupo de trabajo del consumidor, si es necesario. Para obtener información sobre cómo crear un clúster, consulte Creación de un clúster.
2. Para enumerar las bases de datos del clúster o grupo de trabajo del consumidor de AWS Glue Data Catalog a las que tienen acceso los usuarios, ejecute el comando SHOW DATABASES.

   ```
   SHOW DATABASES FROM DATA CATALOG [ACCOUNT <account-id>,<account-id2>] [LIKE
    <expression>]
   ```

   Al hacerlo, se muestran los recursos que están disponibles en el catálogo de datos, como el ARN de la base de datos de AWS Glue, el nombre de la base de datos y la información sobre el recurso compartido de datos.
3. Con el ARN de la base de datos de AWS Glue obtenido mediante SHOW DATABASES, cree una base de datos local en el clúster o grupo de trabajo del consumidor. Para obtener más información, consulte CREATE DATABASE.

   ```
   CREATE DATABASE lf_db FROM ARN <lake-formation-database-ARN> WITH [NO] DATA CATALOG
    SCHEMA [<schema>];
   ```
4. Conceda acceso a las bases de datos y las referencias de esquema creadas a partir de los recursos compartidos de datos a usuarios y roles del clúster o grupo de trabajo del consumidor, según sea necesario. Para obtener más información, consulte GRANT o REVOKE. Tenga en cuenta que los usuarios creados a partir del comando CREATE USER no pueden acceder a los objetos del recurso compartido de datos que se hayan compartido con Lake Formation. Solo los usuarios con acceso a Redshift y Lake Formation pueden acceder a los recursos compartidos de datos que se hayan compartido con Lake Formation.

   ```
   GRANT USAGE ON DATABASE sales_db TO IAM:Bob;
   ```

   Como administrador de un clúster o grupo de trabajo del consumidor, solo puede asignar permisos para toda la base de datos creada a partir del recurso compartido de datos a los

usuarios y roles. En algunos casos, necesita controles pormenorizados sobre un subconjunto de los objetos de base de datos creados a partir del datashare.

También puede crear vistas de enlace de tiempo de ejecución encima de los objetos compartidos y utilizarlas para asignar permisos pormenorizados. También puede considerar la posibilidad de indicar a los clústeres del productor que creen recursos compartidos de datos adicionales en su nombre con el nivel de detalle requerido. Puede crear tantas referencias de esquema para la base de datos creada a partir del datashare como quiera.

5. Los usuarios de base de datos pueden usar las vistas SVV_EXTERNAL_TABLES y SVV_EXTERNAL_COLUMNS para buscar todas las tablas o columnas compartidas en la base de datos de AWS Glue.

```
SELECT * from svv_external_tables WHERE redshift_database_name = 'lf_db';

SELECT * from svv_external_columns WHERE redshift_database_name = 'lf_db';
```

6. Consulte los datos de los objetos compartidos que se encuentran en los datashares.

Los usuarios y los roles que tengan permisos en las bases de datos del consumidor y los esquemas de los clústeres o grupos de trabajo del consumidor pueden explorar y navegar por los metadatos de cualquier objeto compartido. También pueden explorar y navegar por los objetos locales de un clúster o grupo de trabajo del consumidor. Para ello, pueden utilizar controladores JDBC u ODBC, o vistas las SVV_ALL y SVV_EXTERNAL.

```
SELECT * FROM lf_db.schema.table;
```

Solo se pueden utilizar las instrucciones SELECT en los objetos compartidos. No obstante, puede crear tablas en el clúster consumidor consultando los datos de los objetos compartidos de una base de datos local diferente.

```
// Connect to a local cluster database

// Create a view on shared objects and access it.

CREATE VIEW sales_data
AS SELECT *
FROM sales_db.public.tickit_sales_redshift
WITH NO SCHEMA BINDING;
```

```
SELECT * FROM sales_data;
```

## Introducción a las escrituras en varios almacenamientos mediante el uso compartido de datos en Amazon Redshift

Puede compartir objetos de base de datos para lecturas y escrituras entre distintos clústeres de Amazon Redshift o grupos de trabajo de Amazon Redshift sin servidor dentro de la misma Cuenta de AWS, entre cuentas y entre regiones. En este tema, se explican los procedimientos para configurar el uso compartido de datos que incluye permisos de escritura. Puede conceder permisos como SELECT, INSERT y UPDATE para distintas tablas y USAGE y CREATE para esquemas.

Los datos están disponibles en todos los almacenamientos en cuanto se realiza una transacción de escritura. Los administradores de cuentas de productores pueden determinar si los espacios de nombres o regiones específicos tienen acceso de solo lectura, lectura y escritura o de cualquier otro tipo a los datos. Los procedimientos suponen que está trabajando en una base de datos de un clúster aprovisionado o de un grupo de trabajo de Amazon Redshift sin servidor.

Con Amazon Redshift, puede administrar el uso compartido de datos con escrituras mediante la consola o la interfaz de SQL para controlar el acceso y gobernar los datos en todos los clústeres de Amazon Redshift y cuentas de AWS. En las siguientes secciones se proporcionan instrucciones paso a paso sobre la configuración y administración del uso compartido de datos con escrituras mediante Amazon Redshift.

Para obtener una lista de las regiones en las que está disponible el uso compartido de datos, consulte Regiones de AWS en las que está disponible el uso compartido de datos. Para conocer las consideraciones y limitaciones relativas a las escrituras, consulte Consideraciones sobre el uso compartido de datos en Amazon Redshift.

> **Note**
>
> Las escrituras en varios almacenamientos de Amazon Redshift mediante el uso compartido de datos solo se admiten en el parche 186 de Amazon Redshift para los clústeres aprovisionados en la versión 1.0.78881 actual o posteriores y para los grupos de trabajo de Amazon Redshift sin servidor en la versión 1.0.78890 o posteriores.

Temas

- Conexión a una base de datos en Amazon Redshift
- Acciones de productor para nuevos recursos compartidos de datos en Amazon Redshift
- Acciones del consumidor para nuevos recursos compartidos de datos en Amazon Redshift
- Acciones de productores para recursos compartidos de datos existentes en Amazon Redshift
- Acciones de consumidores para recursos compartidos de datos existentes en Amazon Redshift


## Conexión a una base de datos en Amazon Redshift

Con Amazon Redshift, puede establecer una conexión con el clúster de almacenamiento de datos y ejecutar consultas SQL, cargar datos o realizar tareas administrativas. La conexión a una base de datos se refiere al proceso de creación de un canal seguro entre una aplicación o herramienta cliente y el clúster de Amazon Redshift. En las siguientes secciones se proporcionan instrucciones paso sobre cómo conectarse a una base de datos de Amazon Redshift.

Console

Conéctese a una base de datos para ver bases de datos y objetos dentro de bases de datos o para ver recursos compartidos de datos en el almacenamiento de datos de Amazon Redshift. Las credenciales de usuario utilizadas para conectarse a una base de datos especificada deben tener los permisos necesarios para ver todos los recursos compartidos de datos.

Si no hay ninguna conexión local, realice alguna de las siguientes operaciones:

- Si es administrador de productores, vaya a la pestaña Clústeres para ver los clústeres aprovisionados o a la pestaña Configuración de espacios de nombres para ver los puntos de conexión sin servidor. Seleccione el clúster o espacio de nombres correspondiente en la lista.
- En la página de detalles del clúster o espacio de nombres, en la pestaña Recursos compartidos de datos, seleccione Conectar con la base de datos y realice una de las siguientes acciones:
  - En la sección Recursos compartidos de datos de otros espacios de nombres y Cuentas de AWS, consulte los recursos compartidos de datos de otros clústeres, espacios de nombres o cuentas.
  - En la sección Recursos compartidos de datos creados en mi clúster, consulte los recursos compartidos de datos de su clúster.
- En la ventana Connect to database (Conectarse a la base de datos), realice alguna de las siguientes operaciones:

- Si elige Create a new connection (Crear una nueva conexión), elija AWS Secrets Manager para usar un secreto almacenado con objeto de autenticar el acceso a la conexión.

  O bien, elija Temporary credentials (Credenciales temporales) para utilizar credenciales de base de datos y con ellas autenticar el acceso de la conexión. Especifique valores para Database name (Nombre de base de datos) y Database user (Usuario de la base de datos).

  Elija Conectar.
- Seleccione Use a recent connection (Utilizar una conexión reciente) para conectarse a otra base de datos para la que tenga los permisos necesarios.

  Amazon Redshift establece la conexión automáticamente.

Una vez establecida la conexión con la base de datos, puede comenzar a crear recursos compartidos de datos, consultarlos o crear bases de datos a partir de ellos.

## Acciones de productor para nuevos recursos compartidos de datos en Amazon Redshift

Con Amazon Redshift, puede compartir datos en directo entre los clústeres de Amazon Redshift o las cuentas de AWS mediante recursos compartidos de datos. Un recurso compartido de datos es un objeto de consumidor-productor que le permite compartir datos en directo del clúster de Amazon Redshift con otros clústeres o cuentas de AWS. La creación de recursos compartidos de datos permite asegurar el uso compartido de datos mientras mantiene el control sobre el acceso y garantiza que los datos permanezcan actualizados. En las siguientes secciones se proporcionan detalles sobre la creación de recursos compartidos de datos y la agregación de objetos de base de datos, como esquemas, tablas y vistas, para compartir datos en directo de forma segura.

Temas


- Creación de un recurso compartido de datos en Amazon Redshift
- Adición de objetos a un recurso compartido de datos en Amazon Redshift
- Adición de consumidores de datos a un recurso compartido de datos en Amazon Redshift
- Autorización de un recurso compartido de datos en Amazon Redshift

## Creación de un recurso compartido de datos en Amazon Redshift

Un recurso compartido de datos es un contenedor lógico de objetos, permisos y consumidores de la base de datos. Los consumidores son clústeres aprovisionados de Amazon Redshift o espacios de nombres de Amazon Redshift sin servidor de su cuenta y de otras Cuentas de AWS. Cada recurso compartido de datos está asociado a la base de datos en la que se creó y solo se pueden agregar objetos de esa base de datos. Como administrador de productores, puede crear recursos compartidos de datos en la consola y mediante SQL utilizando uno de los siguientes procedimientos.

### Console

En la consola, puede crear recursos compartidos de datos desde las pestañas Recursos compartidos de datos de la página de detalles del clúster o del espacio de nombres. Una vez creado el recurso compartido de datos, puede crear bases de datos a partir de él en un consumidor como administrador de consumidores.

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. En el menú de navegación, elija Clusters (Clústeres) y, a continuación, elija su clúster. Se abrirá la página de detalles del clúster.
3. En la página de detalles del clúster o del espacio de nombres, en la pestaña Recursos compartidos de datos, en la sección Recursos compartidos de datos, conéctese a una base de datos si aún no está conectado. En la sección Recursos compartidos de datos creados en mi cuenta, elija Crear un recurso compartido de datos. Se abrirá la página Crear un recurso compartido de datos.
4. Elija Create datashare (Crear datashare). Solo puede crear un datashare a partir de una base de datos local. Si no se ha conectado a la base de datos, aparecerá la página Conectarse a la base de datos. Siga los pasos de Conexión a una base de datos para conectarse a una base de datos. Si hay una conexión reciente, se abrirá la página Crear un recurso compartido de datos.
5. En la sección Datashare information (Información sobre los recursos compartidos de datos), elija una de las siguientes opciones:
   - Elija Recurso compartido de datos para crear recursos compartidos de datos con el objetivo de compartir datos con fines de lectura o escritura entre diferentes almacenamientos de datos de Amazon Redshift (clústeres aprovisionados o puntos de conexión sin servidor) o bien en la misma Cuenta de AWS o en diferentes Cuentas de AWS.

- Elija AWS Data Exchange datashare (Recurso compartido de datos de ) para crear recursos compartidos de datos con objeto de obtener licencias para los datos a través de AWS Data Exchange.

6. Especifique valores para Datashare name (Nombre del recurso compartido de datos), Database name (Nombre de la base de datos) y Publicly accessible (De acceso público). Cuando cambie el nombre de la base de datos, establezca una nueva conexión a la base de datos.
7. Añada objetos a su recurso compartido de datos mediante las secciones Permisos con ámbito o Permisos directos. Para añadir objetos a un recurso compartido de datos, consulte Creación de un recurso compartido de datos en Amazon Redshift.
8. En la sección Consumidores de datos, puede optar por publicar en Amazon Redshift o en el AWS Glue Data Catalog, lo que inicia el proceso para compartir datos con Lake Formation. Al publicar el recurso compartido de datos en Amazon Redshift, los datos se comparten con otro espacio de nombres o cuenta de Amazon Redshift que actúa como consumidor.

   > **Note**
   >
   > Una vez creado el recurso compartido de datos, no puede editar la configuración para realizar la publicación en la otra opción.

9. Elija Create datashare (Crear datashare).

SQL

El siguiente comando crea un recurso compartido de datos:

```
CREATE DATASHARE salesshare;
```

En el momento de crear el recurso compartido de datos, cada recurso compartido de datos se asocia a una base de datos. Solo los objetos de esa base de datos se pueden compartir en ese datashare. Se pueden crear varios datashares en la misma base de datos con el mismo nivel de detalle de los objetos o uno diferente. No hay límite en cuanto a la cantidad de datashares que puede crear un clúster. También puede utilizar la consola de Amazon Redshift para crear datashares. Para obtener más información, consulte CREATE DATASHARE.

También puede controlar las restricciones de seguridad del recurso compartido de datos durante su creación. En el siguiente ejemplo, se muestra que el consumidor que cuenta con acceso de IP pública tiene permitido leer el recurso compartido de datos.

```
CREATE DATASHARE my_datashare [PUBLICACCESSIBLE = TRUE];
```

Si se establece PUBLICACCESSIBLE como TRUE, los consumidores pueden consultar su recurso compartido de datos desde clústeres de acceso público y grupos de trabajo aprovisionados. Omita esta opción o defínala explícitamente como falsa si no quiere permitirla.

Puede modificar las propiedades referidas al tipo de consumidores después de crear un recurso compartido de datos. Por ejemplo, puede definir que los clústeres que deseen consumir datos de un determinado datashare no puedan ser de acceso público. Las consultas de los clústeres consumidores que no cumplan las restricciones de seguridad especificadas en el recurso compartido de datos se rechazarán en el tiempo de ejecución de la consulta. Para obtener más información, consulte ALTER DATASHARE.

## Adición de objetos a un recurso compartido de datos en Amazon Redshift

Puede añadir objetos de base de datos de varios tipos en la consola y mediante SQL utilizando uno de los siguientes procedimientos.

### Console

Puede añadir objetos al recurso compartido de datos utilizando las secciones Permisos con ámbito o Permisos directos. Seleccione Conceder permisos con ámbito o Conceder permisos directos para añadir objetos. Seleccione el botón Añadir para añadir objetos. Aparecerá un cuadro de diálogo. Siga estos pasos:

1. Si selecciona Conceder permisos con ámbito, aparece la página Conceder permisos con ámbito, donde puede conceder permisos con ámbito en la base de datos o el esquema. Los recursos compartidos de datos con permisos con ámbito tienen los permisos especificados en todos los objetos actuales y futuros de la base de datos o del esquema. Para obtener más información, consulte Permisos acotados.

    1. A continuación, seleccione Permisos con ámbito de base de datos para conceder permisos con ámbito en la base de datos. Al conceder permisos con ámbito, estos se aplican a la base de datos actual al crear el recurso compartido de datos. Estos permisos

no se pueden conceder a objetos individuales y se aplican tanto a los objetos existentes como a los nuevos (esquemas, tablas, vistas y UDF).

2. Seleccione los permisos con ámbito para los esquemas, tablas, vistas o funciones definidas por el usuario. Esto significa que todos los objetos de la base de datos tienen los permisos seleccionados que se han concedido a los consumidores. Seleccione Conceder para completar la concesión de los permisos con ámbito de la base de datos.
3. A continuación, seleccione Permisos con ámbito de esquema para conceder permisos con ámbito en el esquema. Al conceder permisos con ámbito de esquema, todos los objetos que se añaden al esquema tienen los permisos de recurso compartido de datos especificados.
4. Seleccione en el menú desplegable el esquema que desea añadir al recurso compartido de datos. Solo puede seleccionar un esquema a la vez. A continuación, seleccione los permisos directos que desee conceder en el esquema seleccionado.
5. Seleccione permisos con ámbito para sus objetos de esquema, como tablas, vistas y funciones definidas por el usuario. Los permisos se conceden a todos los objetos correspondientes del esquema. Pueden ser objetos existentes o que se añadan en el futuro. Cuando se aplica, el permiso no se puede eliminar de un objeto sin revocar los permisos con ámbito.
6. Seleccione Conceder para completar la concesión de los permisos con ámbito de esquema.

2. Si selecciona Conceder permisos directos, aparece la página Conceder permisos directos, en la que puede conceder permisos directos a cada objeto, como un esquema, una tabla, una vista o una función definida por el usuario. Para conceder permisos directos, primero debe añadir los esquemas pertinentes al recurso compartido de datos.

   1. A continuación, seleccione Conceder permisos directos a los esquemas para aplicar los permisos directos a un esquema específico. Luego, seleccione los permisos de esquema para los objetos de esquema, como tablas, vistas y funciones definidas por el usuario, y seleccione el esquema que desee añadir al recurso compartido de datos. Seleccione Conceder para completar la adición del esquema al recurso compartido de datos.
   2. Una vez que haya añadido un esquema al recurso compartido de datos, puede continuar añadiendo permisos directos para los objetos del esquema. Vuelva a seleccionar Conceder permisos directos. Aparece la página Conceder permisos directos. A continuación, vaya a las pestañas de permisos directos de los objetos del esquema.

3. Seleccione Conceder permisos directos a tablas y vistas para conceder permisos directos de objeto en estos objetos. Seleccione los permisos directos necesarios y los objetos necesarios en la lista. Puede utilizar el campo de búsqueda para buscar objetos de recursos compartidos de datos. Seleccione Conceder para terminar de añadir tablas y vistas al recurso compartido de datos.
4. Seleccione Conceder permisos directos a funciones definidas por el usuario para conceder permisos directos de objeto a funciones definidas por el usuario. Seleccione los permisos directos necesarios y los objetos necesarios en la lista. Puede utilizar el campo de búsqueda para buscar objetos de recursos compartidos de datos. Seleccione Conceder para terminar de añadir funciones definidas por el usuario al recurso compartido de datos.

3. También puede elegir si desea Añadir objetos futuros. Si decide incluir los objetos compartidos de datos agregados al esquema, significa que los objetos agregados al esquema se agregan al recurso compartido de datos automáticamente.
4. Seleccione Agregar para completar la sección y añadir los objetos. Estos se enumeran bajo Objetos del recurso compartido de datos.
5. Tras añadir objetos, puede seleccionar objetos individuales y editar sus permisos. Si selecciona un esquema, aparece un cuadro de diálogo en el que se le pregunta si desea añadir Permisos acotados. Esto hace que cada objeto existente o añadido al esquema tenga un conjunto de permisos preseleccionados, adecuados para el tipo de objeto. Por ejemplo, el administrador puede configurar que todas las tablas añadidas tengan los permisos SELECT y UPDATE.
6. Todos los objetos de recursos compartidos de datos se muestran en las secciones Permisos con ámbito o Permisos directos.
7. En la sección Consumidores de datos, puede agregar espacios de nombres o cuentas de AWS como consumidores del recurso compartido de datos.
8. Elija Crear un recurso compartido de datos para guardar los cambios.

Después de crear el recurso compartido de datos, este aparece en la lista Recursos compartidos de datos creados en mi espacio de nombres. Si elige un recurso compartido de datos de la lista, puede ver sus consumidores, sus objetos y otras propiedades.

SQL

Con SQL, el propietario del recurso compartido de datos debe conceder USAGE en los esquemas que quiera añadir al recurso compartido de datos. GRANT se utiliza para conceder varias

acciones en el esquema, incluidas CREATE y USAGE. Los esquemas contienen objetos compartidos:

```
CREATE SCHEMA myshared_schema1;
CREATE SCHEMA myshared_schema2;

GRANT USAGE ON SCHEMA myshared_schema1 TO DATASHARE my_datashare;
GRANT CREATE, USAGE ON SCHEMA myshared_schema2 TO DATASHARE my_datashare;
```

Como alternativa, el administrador puede seguir ejecutando los comandos ALTER para añadir un esquema al recurso compartido de datos. Solo se conceden permisos de USAGE cuando se agrega un esquema de esta manera.

```
ALTER DATASHARE my_datashare ADD SCHEMA myshared_schema1;
```

Una vez que el administrador ha agregado los esquemas, puede conceder permisos de recurso compartido de datos sobre los objetos del esquema. Pueden ser permisos de lectura y escritura. En el ejemplo GRANT ALL se muestra cómo conceder todos los permisos.

```
GRANT SELECT, INSERT ON TABLE myshared_schema1.table1, myshared_schema1.table2,
 myshared_schema2.table1
TO DATASHARE my_datashare;

GRANT ALL ON TABLE myshared_schema1.table4 TO DATASHARE my_datashare;
```

Puede seguir ejecutando comandos como ALTER DATASHARE para añadir tablas. Cuando lo haga, solo se concederán los permisos SELECT a los objetos añadidos.

```
ALTER DATASHARE my_datashare ADD TABLE myshared_schema1.table1,
 myshared_schema1.table2, myshared_schema2.table1;
```

Puede conceder permisos con ámbito a un recurso compartido de datos en todos los objetos de un tipo en una base de datos o esquema. Los recursos compartidos de datos con permisos con ámbito tienen los permisos especificados en todos los objetos actuales y futuros de la base de datos o del esquema.

Puede ver el alcance de los permisos limitados en el nivel de base de datos en SVV_DATABASE_PRIVILEGES. Puede ver el alcance de los permisos limitados en el nivel de esquema en SVV_SCHEMA_PRIVILEGES.

Esta es la sintaxis para conceder permisos con ámbito a recursos compartidos de datos. Para obtener más información sobre los permisos acotados, consulte Permisos acotados.

```
GRANT { CREATE | USAGE | ALTER | DROP } [,...] | ALL [ PRIVILEGES ] }FOR SCHEMAS IN
DATABASE db_name
TO DATASHARE { datashare_name}

GRANT { { SELECT | INSERT | UPDATE | DELETE | DROP | ALTER | TRUNCATE | REFERENCES }
 [, ...] } | ALL [PRIVILEGES] } }FOR TABLES IN
{SCHEMA schema_name [DATABASE db_name ] | DATABASE db_name }
TO DATASHARE { datashare_name}

GRANT { EXECUTE | ALL [ PRIVILEGES ] }FOR FUNCTIONS IN
{SCHEMA schema_name [DATABASE db_name ] | DATABASE db_name }
TO DATASHARE { datashare_name}
```

Adición de consumidores de datos a un recurso compartido de datos en Amazon Redshift

Puede añadir uno o más consumidores de datos a los recursos compartidos de datos en la consola o mediante SQL. Los consumidores de datos pueden ser espacios de nombres que identifican de forma exclusiva los clústeres de Amazon Redshift o las Cuentas de AWS.

Console

Debe elegir explícitamente si desea habilitar o no la opción de compartir su recurso compartido de datos con los clústeres de acceso público.

- Seleccione Agregar espacios de nombres al recurso compartido de datos. El espacio de nombres es el identificador global único (GUID) del clúster de Amazon Redshift.
- Elija Add Cuentas de AWS (Agregar Cuentas de AWS) al recurso compartido de datos. Las Cuentas de AWS especificadas deben tener permisos de acceso al recurso compartido de datos.

SQL

Con SQL, el administrador concede el uso del recurso compartido de datos a un espacio de nombres específico de la cuenta. Encontrará el ID del espacio de nombres como parte del ARN en la página de detalles del clúster, en la página de detalles del espacio de nombres de Amazon

Redshift sin servidor o al ejecutar el comando `SELECT current_namespace;`. Para obtener más información, consulte CURRENT_NAMESPACE.

```
GRANT USAGE ON DATASHARE my_datashare TO NAMESPACE '86b5169f-012a-234b-9fbb-
e2e24359e9a8';
```

A continuación, se ofrece un ejemplo de cómo conceder el uso de un recurso compartido de datos a una Cuenta de AWS.

```
GRANT USAGE ON DATASHARE salesshare TO ACCOUNT '123456789012';
```

A continuación, se ofrece un ejemplo de cómo conceder el uso de un recurso compartido de datos a una cuenta de Lake Formation.

```
GRANT USAGE ON DATASHARE salesshare TO ACCOUNT '123456789012' VIA DATA CATALOG;
```

## Autorización de un recurso compartido de datos en Amazon Redshift

Con Amazon Redshift, puede controlar el acceso a los recursos compartidos de datos mediante la autorización de consumidores específicos. Los recursos compartidos de datos le permiten compartir datos en directo entre clústeres de Amazon Redshift en las mismas cuentas de AWS o diferentes, lo que proporciona una forma sencilla de distribuir y consumir datos de análisis. En esta sección se proporcionan instrucciones paso a paso para autorizar y revocar el acceso del consumidor a los recursos compartidos de datos en Amazon Redshift.

> **Note**
>
> Si va a añadir un espacio de nombres como consumidor de datos, no tiene que realizar el proceso de autorización. Para autorizar un recurso compartido de datos, debe haber al menos un consumidor de datos añadido al recurso compartido de datos.

### Console

Como administrador de productores en la consola, puede elegir a qué consumidores de datos autorizará para que accedan a los recursos compartidos de datos o a qué consumidores de datos retirará la autorización. Los consumidores de datos autorizados reciben notificaciones para tomar

medidas en los datashares. Si va a añadir un espacio de nombres como consumidor de datos, no tiene que realizar el proceso de autorización.

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. En el menú de navegación, elija Datashares (Recursos compartidos de datos). Aquí puede ver una lista llamada Consumidores de recursos compartidos de datos. Elija uno o más recursos compartidos de datos que desee autorizar. A continuación elija Authorize.
3. Aparece el cuadro de diálogo Autorizar la cuenta. Puede elegir entre un par de tipos de autorización.

   - Solo lectura en [nombre del clúster o nombre del grupo de trabajo]: esto significa que el consumidor no dispone de permisos de escritura, aunque el creador del recurso compartido de datos los haya concedido.
   - Lectura y escritura en [nombre del clúster o nombre del grupo de trabajo]: esto significa que el consumidor dispone de todos los permisos otorgados por el creador, incluso de escritura.
4. Seleccione Save.

También puede autorizar a AWS Data Exchange como consumidor.

1. Si ha elegido Publicar en AWS Glue Data Catalog al crear el recurso compartido de datos, solo puede conceder la autorización del recurso compartido de datos a una cuenta de Lake Formation.

   Para el recurso compartido de datos de AWS Data Exchange, solo puede autorizar un recurso compartido de datos a la vez.

   Cuando autoriza un recuso compartido de datos de AWS Data Exchange, comparte el recurso compartido de datos con el servicio AWS Data Exchange y permite que AWS Data Exchange administre el acceso al recurso compartido de datos en su nombre. Para que los consumidores obtengan acceso, AWS Data Exchange agrega cuentas consumidoras como consumidores de datos al recurso compartido de datos de AWS Data Exchange cuando se suscriben a los productos. AWS Data Exchange no tiene acceso de lectura al recurso compartido de datos.
2. Seleccione Save.

Una vez autorizados los consumidores de datos, ellos pueden acceder a los objetos de los datashares y crear una base de datos consumidora para consultar los datos.

API

El administrador de seguridad del productor determina lo siguiente:

- Si otra cuenta puede acceder o no al recurso compartido de datos.
- Si una cuenta tiene acceso al recurso compartido de datos, sin importar si dicha cuenta tiene o no permisos de escritura.

Los siguientes permisos de IAM son necesarios para autorizar un recurso compartido de datos:

redshift:AuthorizeDataShare

Puede autorizar el uso y las escrituras mediante una llamada a la CLI o con la API:

```
authorize-data-share
--data-share-arn <value>
--consumer-identifier <value>
[--allow-writes | --no-allow-writes]
```

Para obtener más información acerca del comando, consulte authorize-data-share.

El identificador del consumidor puede ser:

- Un identificador de cuenta de AWS de doce dígitos.
- El ARN del identificador del espacio de nombres.

> **Note**
>
> Los permisos de escritura no se conceden en el paso de autorización. Al autorizar un recurso compartido de datos para las escrituras, la cuenta solo dispondrá de permisos de escritura otorgados por el administrador del recurso compartido de datos. Si un administrador no permite la escritura, los únicos permisos disponibles para el consumidor en concreto son SELECT, USAGE y EXECUTE.

Para cambiar la autorización de un consumidor de recursos compartidos de datos, vuelva a llamar a `authorize-data-share`, pero con un valor diferente. La autorización anterior se sobrescribe

con la nueva autorización. Por lo tanto, si inicialmente autoriza y permite la escritura, pero vuelve a autorizar y especifica `no-allow-writes` o simplemente no especifica un valor, se revocarán los permisos de escritura del consumidor.

## Acciones del consumidor para nuevos recursos compartidos de datos en Amazon Redshift

Con Amazon Redshift, puede consumir recursos compartidos de datos de otras cuentas de AWS, lo que permite la colaboración y el uso compartido de datos entre cuentas. Un recurso compartido de datos es una forma segura de compartir datos en directo entre clústeres de Amazon Redshift, incluso si se encuentran en cuentas de AWS diferentes. En las siguientes secciones se proporcionan pasos detallados para configurar el acceso, crear bases de datos a partir de recursos compartidos de datos, conceder permisos de objeto y consultar datos compartidos.

Temas


- Asociación de un recurso compartido de datos desde otra Cuenta de AWS en Amazon Redshift
- Creación de una base de datos a partir de un recurso compartido de datos de Amazon Redshift
- Concesión de permisos de objeto a usuarios y roles de consumidor en Amazon Redshift
- Consulta de datos en un recurso compartido de datos de Amazon Redshift


### Asociación de un recurso compartido de datos desde otra Cuenta de AWS en Amazon Redshift

Con Amazon Redshift, puede asociar los recursos compartidos de datos que se comparten desde otras Cuentas de AWS, lo que permite compartir datos de forma segura y sin problemas a través de los límites de la organización. Los recursos compartidos de datos son objetos de bases de datos que se pueden compartir y que encapsulan datos de una o más bases de datos de Amazon Redshift. En las siguientes secciones se muestra el proceso de asociación de recursos compartidos de datos.

Console

Como administrador de consumidores, puede asociar uno o más recursos compartidos de datos que se comparten desde otras cuentas a toda su cuenta de AWS o a espacios de nombres específicos de su cuenta.

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.

2. En el menú de navegación, elija Datashares (Recursos compartidos de datos). Se abrirá la página con la lista de datashares. Elija From other accounts (De otras cuentas).
3. En la sección Datashares from other accounts (Recursos compartidos de datos de otras cuentas), elija el recurso compartido de datos que desee asociar y luego elija Associate (Asociar). Cuando se muestre la página Asociar recursos compartidos de datos, elija uno de los siguientes tipos de asociación:

   - Elija Toda la cuenta de AWS para asociar todos los espacios de nombres existentes y futuros de diferentes regiones de AWS de su cuenta de AWS al recurso compartido de datos.
   - Si el recurso compartido de datos está publicado en AWS Glue Data Catalog, solo puede asociarlo a la cuenta de AWS completa.
4. Desde aquí, puede elegir Permisos admitidos. Estas opciones son las siguientes:

   - Solo lectura: si elige solo lectura, los permisos de escritura como UPDATE o INSERT no estarán disponibles para el consumidor, incluso si el productor los ha concedido y autorizado.
   - Lectura y escritura: los usuarios de recursos compartidos de datos del consumidor dispondrán de todos los permisos, tanto de lectura como de escritura, concedidos y autorizados por el productor.
5. Como alternativa, elija Regiones de AWS y espacios de nombres de clústeres específicos para asociar una o más regiones de AWS y espacios de nombres específicos al recurso compartido de datos. Elija Agregar región para añadir regiones de AWS y espacios de nombres específicos al recurso compartido de datos. Aparecerá la página Add AWS Region (Agregar región de AWS).
6. Elija una AWS Region (Región de AWS).
7. Realice una de las siguientes acciones:

   - Elija Agregar todos los espacios de nombres de clúster para añadir todos los espacios de nombres de clúster existentes y futuros de esta región al recurso compartido de datos.
   - Elija Agregar espacios de nombres específicos para añadir uno o más espacios de nombres específicos de esta región al recurso compartido de datos.
   - Elija uno o más espacios de nombres y, a continuación, Agregar región de AWS.
8. Elija Asociar.

El productor puede volver atrás y cambiar la configuración de una autorización, lo que puede afectar a la configuración de la asociación de los consumidores.

Si va a asociar el recurso compartido de datos a una cuenta de Lake Formation, vaya a la consola de Lake Formation para crear una base de datos y, a continuación, defina los permisos sobre la base de datos. Para obtener más información, consulte Configuración de permisos para los recursos compartidos de datos de Amazon Redshift en la Guía para desarrolladores de AWS Lake Formation. Una vez creada una base de datos de AWS Glue o una base de datos federada, puede utilizar el editor de consultas v2 o cualquier cliente SQL preferido con su clúster consumidor para consultar los datos.

Después de asociar los recursos compartidos de datos, se vuelven disponibles.

> **Note**
>
> También puede cambiar la asociación del datashare en cualquier momento. Cuando se cambia la asociación de regiones de AWS y espacios de nombres específicos a toda la cuenta de AWS, Amazon Redshift sobrescribe la información de las regiones y los espacios de nombres específicos con la información de la cuenta de AWS. A continuación, todas las regiones de AWS y los espacios de nombres en la cuenta de Cuenta de AWS tienen acceso al recurso compartido de datos.

API

> **Note**
>
> Los pasos de esta sección se llevan a cabo después de que el administrador de productores autorice acciones específicas en los objetos de base de datos compartidos y, si el recurso compartido de datos se comparte con otra cuenta, el administrador de seguridad de productores autorice el acceso.

El administrador de seguridad del consumidor determina lo siguiente:

- Si todos los espacios de nombres de una cuenta, los espacios de nombres de regiones específicas de la cuenta o los espacios de nombres específicos tienen acceso al recurso compartido de datos.

- Si los espacios de nombres tienen acceso al recurso compartido de datos, sin importar si tienen o no permisos de escritura.

El administrador de seguridad de consumidores puede asociar el recurso compartido de datos con el siguiente comando:

```
associate-data-share-consumer
--data-share-arn <value>
--consumer-identifier <value>
[--allow-writes | --no-allow-writes]
```

Para obtener más información acerca del comando, consulte associate-data-share-consumer.

El administrador de seguridad del consumidor debe establecer explícitamente el valor `allow-writes` en true al asociar un recurso compartido de datos a un espacio de nombres para permitir el uso de los comandos INSERT y UPDATE. Si no lo hace, los usuarios solo pueden realizar operaciones de lectura, como los privilegios SELECT, USAGE o EXECUTE.

Para cambiar la asociación de un espacio de nombres para un recurso compartido de datos, vuelva a llamar a `associate-data-share-consumer` con un valor diferente. La asociación antigua se sobrescribirá con la nueva, por lo que si asocia y establece `allow-writes` al principio, pero asocia y especifica `no-allow-writes`, o simplemente no especifica ningún valor, se revocarán los permisos de escritura del consumidor.

### Creación de una base de datos a partir de un recurso compartido de datos de Amazon Redshift

Con Amazon Redshift, puede utilizar un recurso compartido de datos para crear una base de datos y, a continuación, consultar datos en recursos compartidos de datos de clústeres de productor para obtener acceso seguro a datos en directo sin copiarlos ni transferirlos. Los siguientes pasos cubren los detalles de la configuración de una base de datos en el entorno de Amazon Redshift.

Console

Para poder consultar los datos del recurso compartido de datos, debe crear una base de datos a partir de un recurso compartido de datos. Solo puede crear una base de datos a partir de un datashare especificado.

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.

2. En el menú de navegación, elija Clusters (Clústeres) y, a continuación, elija su clúster. Se abrirá la página de detalles del clúster.
3. Seleccione Datashares. Aparecerá la lista de datashares.
4. En la sección Datashares from other clusters (Datashares de otros clústeres), elija Connect to database (Conectarse a la base de datos). Para obtener más información, consulte Conexión a una base de datos.
5. Seleccione el datashare a partir del cual desea crear bases de datos y, a continuación, elija Create database from datashare (Crear base de datos a partir del datashare). Se abrirá la página Create database from datashare (Crear base de datos a partir del datashare).
6. En Database name (Nombre de base de datos), especifique un nombre para la base de datos. El nombre de la base de datos debe tener entre 1 y 64 caracteres alfanuméricos (solo letras minúsculas) y no puede ser una palabra reservada.
7. Seleccione Crear.

Una vez creada la base de datos, puede consultar los datos de la base de datos o realizar operaciones de escritura, si el administrador del consumidor las ha concedido, autorizado y asociado.

API

Para compartir datos con fines de lectura como administrador de consumidores, realice los siguientes pasos.

1. Enumere los datashares que se ponen a su disposición y vea su contenido. Para obtener más información, consulte DESC DATASHARE y SHOW DATASHARES.

   En el siguiente ejemplo, se muestra la información de los datashares entrantes de un espacio de nombres productor específico. Cuando ejecuta DESC DATASHARE como administrador de consumidores, debe especificar la opción NAMESPACE para ver los recursos compartidos de datos entrantes.

```
DESC DATASHARE salesshare OF NAMESPACE '13b8833d-17c6-4f16-8fe4-1a018f5ed00d';


 producer_account  |          producer_namespace          | share_type |
 share_name | object_type |           object_name           |   include_new
-------------------+--------------------------------------+------------
+------------+-------------+---------------------------------+-----------------
```

```
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    |
 salesshare | table       | public.tickit_users_redshift    |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    |
 salesshare | table       | public.tickit_venue_redshift    |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    |
 salesshare | table       | public.tickit_category_redshift |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    |
 salesshare | table       | public.tickit_date_redshift     |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    |
 salesshare | table       | public.tickit_event_redshift    |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    |
 salesshare | table       | public.tickit_listing_redshift  |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    |
 salesshare | table       | public.tickit_sales_redshift    |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    |
 salesshare | schema      | public                          |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    |
 salesshare | view        | public.sales_data_summary_view  |
```

Solo los superusuarios del clúster pueden hacer esto. También puede utilizar SVV_DATASHARES para ver los datashares y SVV_DATASHARE_OBJECTS para ver los objetos que se encuentran dentro del datashare.

En el siguiente ejemplo, se muestran los datashares entrantes de un clúster consumidor.

```
SHOW DATASHARES LIKE 'sales%';


 share_name | share_owner | source_database | consumer_database | share_type
 | createdate | is_publicaccessible | share_acl | producer_account |
 producer_namespace
------------+-------------+-----------------+-------------------+------------
+------------+---------------------+-----------+-----------------
+--------------------------------------
 salesshare |             |                 |                   | INBOUND
    |            |         t           |           |   123456789012   |
 13b8833d-17c6-4f16-8fe4-1a018f5ed00d
```

2. Como superusuario de bases de datos, puede crear bases de datos locales que hacen referencia a los recursos compartidos de datos. Para obtener más información, consulte CREATE DATABASE.

```
CREATE DATABASE sales_db FROM DATASHARE salesshare OF NAMESPACE
 '13b8833d-17c6-4f16-8fe4-1a018f5ed00d';
```

Si desea tener un control más detallado del acceso a los objetos de la base de datos local, utilice la cláusula WITH PERMISSIONS al crear la base de datos. De este modo, podrá conceder permisos en el nivel de objeto para los objetos de la base de datos en el paso 4.

```
CREATE DATABASE sales_db WITH PERMISSIONS FROM DATASHARE salesshare OF NAMESPACE
 '13b8833d-17c6-4f16-8fe4-1a018f5ed00d';
```

Puede ver las bases de datos que creó a partir del datashare consultando la vista SVV_REDSHIFT_DATABASES. Puede conectarse a estas bases de datos directamente o puede conectarse a una base de datos local en el clúster del consumidor y realizar una consulta entre bases de datos para consultar los datos de las bases de datos de recursos compartidos de datos.

> **Note**
>
> No puede crear un datashare encima de los objetos de base de datos creados a partir de un datashare existente. No obstante, puede copiar los datos en una tabla diferente en el clúster consumidor, llevar a cabo el procesamiento que sea necesario y, a continuación, compartir los objetos nuevos que se hayan creado.

También puede utilizar la consola de Amazon Redshift para crear bases de datos a partir de datashares. Para obtener más información, consulte Creación de bases de datos a partir de datashares.

Concesión de permisos de objeto a usuarios y roles de consumidor en Amazon Redshift

Como administrador de consumidores, puede conceder permisos a los usuarios y roles de consumidor en el objeto siguiendo estos pasos.

SQL

Si ha creado su base de datos sin WITH PERMISSIONS, solo podrá asignar permisos a toda la base de datos creada a partir del recurso compartido de datos a los usuarios y los roles.

```
GRANT USAGE ON DATABASE sales_db TO Bob;
```

```
GRANT USAGE ON SCHEMA sales_schema TO ROLE Analyst_role;
```

También puede crear vistas de enlace de tiempo de ejecución encima de los objetos compartidos y utilizarlas para asignar permisos pormenorizados. También puede considerar la posibilidad de indicar a los clústeres productores que creen datashares adicionales en su nombre con el nivel de detalle requerido.

Si ha creado la base de datos con WITH PERMISSIONS, debe asignar permisos de objeto para los objetos de la base de datos compartida. Un usuario que solo tenga el permiso USAGE no podrá acceder a ningún objeto de una base de datos creada con WITH PERMISSIONS hasta que se le concedan permisos adicionales en el nivel de objeto.

```
GRANT SELECT ON sales_db.public.tickit_sales_redshift to Bob;
```

Para obtener más información sobre cómo conceder permisos con escrituras de varios almacenamientos, consulte Administración de permisos para un recurso compartido de datos en Amazon Redshift.

Consulta de datos en un recurso compartido de datos de Amazon Redshift

Con Amazon Redshift, puede consultar datos entre recursos compartidos de datos de clústeres de productores para acceder de forma segura a los datos en directo sin necesidad de copiarlos ni transferirlos. En la siguiente sección, se explica la consulta de los recursos compartidos de datos en el entorno de Amazon Redshift.

Los usuarios y los roles que tengan permisos en las bases de datos del consumidor y los esquemas de los clústeres del consumidor pueden explorar y navegar por los metadatos de cualquier objeto compartido. También pueden navegar por los objetos locales de un clúster consumidor y explorarlos. Para ello, utilizan controladores JDBC u ODBC, o vistas SVV_ALL y SVV_REDSHIFT.

Los clústeres productores pueden tener muchos esquemas en la base de datos y muchas tablas y vistas dentro de cada esquema. Los usuarios del lado del consumidor solo pueden ver el subconjunto de objetos que se han puesto a disposición a través del datashare. Estos usuarios no pueden ver todos los metadatos del clúster productor. Este enfoque ayuda a proporcionar un control pormenorizado de la seguridad de los metadatos con el uso compartido de datos.

Continúa conectándose a las bases de datos locales del clúster. No obstante, ahora, también es posible leer de las bases de datos y los esquemas que se crean a partir de los datashares utilizando la notación de tres partes database.schema.table. Puede realizar consultas que abarquen todas las bases de datos visibles para usted. Estas pueden ser bases de datos locales del clúster o bases de datos creadas a partir de los datashares. Los clústeres consumidores no pueden conectarse a las bases de datos que se hayan creado a partir de los datashares.

Puede acceder a los datos mediante la cualificación completa. Para obtener más información, consulte Ejemplos de consultas entre bases de datos.

SQL

```
SELECT * FROM sales_db.public.tickit_sales_redshift ORDER BY 1,2 LIMIT 5;

 salesid | listid | sellerid | buyerid | eventid | dateid | qtysold | pricepaid |
 commission |      saletime
---------+--------+----------+---------+---------+--------+---------+-----------
+------------+---------------------
       1 |      1 |    36861 |   21191 |    7872 |   1875 |       4 |    728.00 |
  109.20 | 2008-02-18 02:36:48
       2 |      4 |     8117 |   11498 |    4337 |   1983 |       2 |     76.00 |
   11.40 | 2008-06-06 05:00:16
       3 |      5 |     1616 |   17433 |    8647 |   1983 |       2 |    350.00 |
   52.50 | 2008-06-06 08:26:17
       4 |      5 |     1616 |   19715 |    8647 |   1986 |       1 |    175.00 |
   26.25 | 2008-06-09 08:38:52
       5 |      6 |    47402 |   14115 |    8240 |   2069 |       2 |    154.00 |
   23.10 | 2008-08-31 09:17:02
```

Solo se pueden utilizar las instrucciones SELECT en los objetos compartidos. No obstante, puede crear tablas en el clúster consumidor consultando los datos de los objetos compartidos de una base de datos local diferente.

## Acciones de productores para recursos compartidos de datos existentes en Amazon Redshift

Con Amazon Redshift, puede administrar los recursos compartidos de datos existentes para controlar el acceso a los datos en un clúster de Amazon Redshift. En las siguientes secciones se proporcionan instrucciones paso a paso sobre la modificación de objetos de recurso compartido de datos, la administración de los permisos de recurso compartido de datos y la actualización de las propiedades

de recurso compartido de datos para controlar y auditar de forma eficaz el acceso a los datos en el entorno de Amazon Redshift.

Temas


- Visualización de un recurso compartido de datos en Amazon Redshift
- Edición de recursos compartidos de datos creados en su cuenta en Amazon Redshift
- Eliminación de la autorización de un recurso compartido de datos en Amazon Redshift
- Eliminación de objetos de recursos compartidos de datos de un recurso compartido de datos en Amazon Redshift
- Eliminación de consumidores de datos de un recurso compartido de datos en Amazon Redshift
- Eliminación de un recurso compartido de datos creado en su cuenta en Amazon Redshift


## Visualización de un recurso compartido de datos en Amazon Redshift

Puede ver los recursos compartidos de datos desde la consola o mediante SQL.

Console

Puede ver los recursos compartidos de datos desde la pestaña Recursos compartidos de datos o Clústeres.

- Use la pestaña Recursos compartidos de datos para ver una lista de los recursos compartidos de datos de su cuenta o de otras cuentas.
  - Para ver los datashares que se han creado en su cuenta, elija In my account (En mi cuenta) y, luego, seleccione el datashare que desee consultar.
  - Para ver los datashares que se comparten desde otras cuentas, elija From other accounts (De otras cuentas) y, luego, seleccione los datashares que desee consultar.
- Use la pestaña Clústeres para ver una lista de los recursos compartidos de datos de su clúster o de otros clústeres.

  Primero, conéctese a una base de datos. A continuación, elija un recurso compartido de datos en la sección Recursos compartidos de datos de otros clústeres o la sección Recursos compartidos de datos creados en mi clúster para ver sus detalles.

SQL

Puede ver una lista de recursos compartidos de datos que se hayan creado en el clúster y examinar su contenido.

En el siguiente ejemplo, se muestra la información de un recurso compartido de datos denominado `salesshare`.

```
DESC DATASHARE salesshare;

 producer_account  |          producer_namespace          | share_type | share_name
 | object_type |           object_name          |   include_new
------------------+--------------------------------------+------------+------------
+-------------+--------------------------------+------------------
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | table       | public.tickit_users_redshift   |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | table       | public.tickit_venue_redshift   |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | table       | public.tickit_category_redshift|
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | table       | public.tickit_date_redshift    |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | table       | public.tickit_event_redshift   |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | table       | public.tickit_listing_redshift |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | table       | public.tickit_sales_redshift   |
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | schema      | public                         |  t
 123456789012      | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | OUTBOUND   | salesshare
 | view        | public.sales_data_summary_view |
```

En el siguiente ejemplo, se muestran los datashares salientes de un clúster productor.

```
SHOW DATASHARES LIKE 'sales%';
```

El resultado tiene un aspecto similar al siguiente.

```
share_name | share_owner  | source_database | consumer_database | share_type |
  createdate      | is_publicaccessible  | share_acl | producer_account |
 producer_namespace
```

```
----------+-------------+----------------+-------------------+-----------
+--------------------+---------------------+----------+-----------------
+--------------------------------------
salesshare |    100       | dev             |                   |  OUTBOUND
 | 2020-12-09 02:27:08 |          True        |           |   123456789012   |
 13b8833d-17c6-4f16-8fe4-1a018f5ed00d
```

Para obtener más información, consulte DESC DATASHARE y SHOW DATASHARES.

También puede utilizar SVV_DATASHARES, SVV_DATASHARE_CONSUMERS y SVV_DATASHARE_OBJECTS para ver los datashares, los objetos dentro del datashare y sus consumidores.

## Edición de recursos compartidos de datos creados en su cuenta en Amazon Redshift

Puede editar un recurso compartido de datos creado en su cuenta utilizando la consola y mediante SQL.

### Console

En la consola, siga estos pasos para conectarse primero a una base de datos y ver la lista de recursos compartidos de datos creados en su cuenta.

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. En el menú de navegación, elija Clusters (Clústeres) y, a continuación, elija su clúster. Se abrirá la página de detalles del clúster.
3. Seleccione Datashares.
4. En la sección Datashares created in my account (Datashares creados en mi cuenta), elija Connect to database (Conectarse a la base de datos).
5. Elija el datashare que desee modificar y, a continuación, seleccione Edit (Editar). Se abrirá la página de detalles del datashare.
6. Realice todos los cambios en la sección Datashare objects (Objetos de datashares) o en la sección Data consumers (Consumidores de datos).
7. Elija Save changes (Guardar cambios). Amazon Redshift actualizará sus datashares con los cambios.

> **Note**
>
> Si eligió publicar su recurso compartido de datos en AWS Glue Data Catalog, no podrá editar la configuración para publicarlo en otras cuentas de Amazon Redshift.

SQL

Utilice ALTER DATASHARE para quitar objetos de los datashares en cualquier parte de este. Utilice el siguiente comando para eliminar un esquema:

```
ALTER DATASHARE salesshare REMOVE SCHEMA PUBLIC;
```

Utilice el siguiente comando para eliminar una tabla:

```
ALTER DATASHARE salesshare REMOVE TABLE public.tickit_sales_redshift;
```

Utilice REVOKE USAGE ON para revocar los permisos sobre el datashare otorgados a determinados consumidores. Revoca los permisos USAGE de los objetos que se encuentran dentro de un recurso compartido de datos y desactiva de forma instantánea el acceso para todos los clústeres consumidores. Las listas de los datashares y las consultas de metadatos, como las listas de bases de datos y tablas, no devuelven los objetos compartidos después de haber revocado el acceso. Revoque el acceso a los datashares desde los espacios de nombres si ya no desea compartir los datos con los consumidores.

```
REVOKE USAGE ON DATASHARE salesshare FROM NAMESPACE
 '13b8833d-17c6-4f16-8fe4-1a018f5ed00d';
```

Si ya no desea compartir los datos con los consumidores, revoque el acceso al recurso compartido de datos de las Cuentas de AWS:

```
REVOKE USAGE ON DATASHARE salesshare FROM ACCOUNT '123456789012';
```

## Eliminación de la autorización de un recurso compartido de datos en Amazon Redshift

Con Amazon Redshift, puede controlar el acceso a los recursos compartidos de datos revocando la autorización para consumidores específicos. En esta sección se proporcionan instrucciones para revocar el acceso de los consumidores a los recursos compartidos de datos en Amazon Redshift.

> **Note**
>
> Para retirar la autorización del recurso compartido de datos, debe haber al menos un consumidor de datos añadido al recurso compartido de datos.

Console

Elija uno o más clústeres consumidores a los que desee retirar la autorización. A continuación, elija Eliminar autorización.

Después de retirar la autorización, los consumidores de datos pierden de inmediato el acceso a los recursos compartidos de datos.

API

El administrador de seguridad del productor determina lo siguiente:

- Si otra cuenta puede acceder o no al recurso compartido de datos.
- Si una cuenta tiene acceso al recurso compartido de datos, sin importar si dicha cuenta tiene o no permisos de escritura.

Los siguientes permisos de IAM son necesarios para eliminar la autorización de un recurso compartido de datos:

redshift:DeauthorizeDataShare

Puede eliminar la autorización de uso y escritura mediante una llamada a la CLI o con la API:

```
deauthorize-data-share
--data-share-arn <value>
--consumer-identifier <value>
```

Para obtener más información acerca del comando, consulte deauthorize-data-share.

## Eliminación de objetos de recursos compartidos de datos de un recurso compartido de datos en Amazon Redshift

Puede eliminar uno o más objetos de un recurso compartido de datos mediante el siguiente procedimiento.

Console

Para eliminar uno o más objetos de un recurso compartido de datos en la consola, siga estos pasos.

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. En el menú de navegación, elija Clusters (Clústeres) y, a continuación, elija su clúster. Se abrirá la página de detalles del clúster.
3. Seleccione Datashares.
4. En la sección Datashares created in my account (Datashares creados en mi cuenta), elija Connect to database (Conectarse a la base de datos). Para obtener más información, consulte Conexión a una base de datos.
5. Elija el datashare que desee modificar y, a continuación, seleccione Edit (Editar). Se abrirá la página de detalles del datashare.
6. Para quitar uno o más objetos del datashare, realice alguna de las siguientes operaciones:
   - Para quitar esquemas del datashare, elija uno o más esquemas. Luego, seleccione Remove (Quitar). Amazon Redshift quitará los esquemas especificados y todos los objetos de los esquemas especificados del datashare.
   - Para quitar tablas y vistas del datashare, elija una o más tablas y vistas. Luego, seleccione Remove (Quitar). También puede elegir Remove by schema (Quitar por esquema) para quitar todas las tablas y las vistas de los esquemas especificados.
   - Para quitar funciones definidas por el usuario del datashare, elija una o más funciones definidas por el usuario. Luego, seleccione Remove (Quitar). También puede elegir Remove by schema (Quitar por esquema) para quitar todas las funciones definidas por el usuario de los esquemas especificados.

SQL

Utilice ALTER DATASHARE para quitar objetos de los datashares en cualquier parte de este. Utilice el siguiente comando para eliminar un esquema:

```
ALTER DATASHARE salesshare REMOVE SCHEMA PUBLIC;
```

Utilice el siguiente comando para eliminar una tabla:

```
ALTER DATASHARE salesshare REMOVE TABLE public.tickit_sales_redshift;
```

### Eliminación de consumidores de datos de un recurso compartido de datos en Amazon Redshift

Puede quitar uno o más consumidores de datos de un datashare. Los consumidores de datos pueden ser espacios de nombres que identifican de forma exclusiva los clústeres de Amazon Redshift o las cuentas de AWS.

Console

Para eliminar uno o más consumidores de datos de un recurso compartido de datos en la consola, elija uno o más consumidores de datos en los ID de espacios de nombres o la cuenta de AWS. Luego, seleccione Eliminar.

Amazon Redshift quita los consumidores de datos especificados del datashare. Los consumidores especificados perderán de inmediato el acceso al datashare.

SQL

Utilice REVOKE USAGE ON para revocar los permisos sobre el datashare otorgados a determinados consumidores. Revoca los permisos USAGE de los objetos que se encuentran dentro de un recurso compartido de datos y desactiva de forma instantánea el acceso para todos los clústeres consumidores. Las listas de los datashares y las consultas de metadatos, como las listas de bases de datos y tablas, no devuelven los objetos compartidos después de haber revocado el acceso. Revoque el acceso a los datashares desde los espacios de nombres si ya no desea compartir los datos con los consumidores.

```
REVOKE USAGE ON DATASHARE salesshare FROM NAMESPACE
 '13b8833d-17c6-4f16-8fe4-1a018f5ed00d';
```

### Eliminación de un recurso compartido de datos creado en su cuenta en Amazon Redshift

Puede eliminar un recurso compartido de datos creado en su cuenta utilizando la consola o mediante SQL.

Console

Para eliminar un recurso compartido de datos creado en su cuenta utilizando la consola, conéctese primero a una base de datos para ver la lista de recursos compartidos de datos creados en su cuenta.

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. En el menú de navegación, elija Clusters (Clústeres) y, a continuación, elija su clúster. Se abrirá la página de detalles del clúster.
3. Seleccione Datashares. Aparecerá la lista de datashares.
4. En la sección Datashares created in my account (Datashares creados en mi cuenta), elija Connect to database (Conectarse a la base de datos).
5. Elija uno o más datashares que desee eliminar y, a continuación, seleccione Delete (Eliminar). Se abrirá la página Delete datashares (Eliminar datashares).

   La eliminación de un recurso compartido de datos con Lake Formation no elimina automáticamente los permisos asociados en Lake Formation. Para eliminarlos, vaya a la consola de Lake Formation.
6. Escriba Delete (Eliminar) para confirmar la eliminación de los datashares especificados.
7. Elija Eliminar.

Después de eliminar los datashares, sus consumidores perderán el acceso a ellos.

SQL

Puede utilizar SQL para eliminar los objetos de recursos compartido de datos en cualquier momento utilizando DROP DATASHARE. Los superusuarios del clúster y los propietarios de los datashares pueden eliminarlos.

En el siguiente ejemplo, se elimina un recurso compartido de datos denominado `salesshare`.

```
DROP DATASHARE salesshare;
```

## Acciones de consumidores para recursos compartidos de datos existentes en Amazon Redshift

Con Amazon Redshift, puede administrar los recursos compartidos de datos existentes para controlar el acceso a los datos en un clúster de Amazon Redshift. En las siguientes secciones se proporcionan instrucciones paso sobre la administración de recursos compartidos de datos en el entorno de Amazon Redshift como administrador de consumidores.

Temas

- Administración de permisos para un recurso compartido de datos en Amazon Redshift
- Eliminación de la asociación de un recurso compartido de datos de los consumidores de datos en una Cuenta de AWS diferente en Amazon Redshift
- Rechazo de un recurso compartido de datos de otra Cuenta de AWS en Amazon Redshift

### Administración de permisos para un recurso compartido de datos en Amazon Redshift

Como administrador de productores, conserva el control de los conjuntos de datos que comparte. Puede agregar objetos nuevos al datashare o quitárselos. También puede otorgar o revocar el acceso a los recursos compartidos de datos en su conjunto para los clústeres consumidores, las cuentas de AWS y las regiones de AWS. Cuando se revocan los permisos, los clústeres consumidores pierden inmediatamente el acceso a los objetos compartidos y dejan de verlos en la lista de recursos compartidos de datos INBOUND en SVV_DATASHARES.

En el siguiente ejemplo, se crea el recurso compartido de datos `salesshare`, se agrega el esquema `public` y se agrega la tabla `public.tickit_sales_redshift` a `salesshare`. También se conceden permisos de uso en `salesshare` al espacio de nombres especificado.

```
CREATE DATASHARE salesshare;

ALTER DATASHARE salesshare ADD SCHEMA public;

ALTER DATASHARE salesshare ADD TABLE public.tickit_sales_redshift;

GRANT USAGE ON DATASHARE salesshare TO NAMESPACE
 '13b8833d-17c6-4f16-8fe4-1a018f5ed00d';
```

Para CREATE DATASHARE, los superusuarios y los propietarios de bases de datos pueden crear datashares. Para obtener más información, consulte CREATE DATASHARE. En el caso de ALTER

DATASHARE, el propietario del datashare que cuenta con los permisos necesarios sobre los objetos del datashare que se van a agregar o quitar, puede modificar el datashare. Para obtener más información, consulta ALTER DATASHARE.

Como el administrador de un productor, cuando elimina un datashare, este deja de aparecer en las listas de los clústeres consumidores. Las referencias a bases de datos y esquema creadas en el clúster consumidor a partir de los datashares eliminados siguen existiendo, aunque no hay ningún objeto en ellas. El administrador de consumidores debe eliminar estas bases de datos de forma manual.

En el caso del consumidor, un administrador de consumidores puede determinar qué usuarios y roles deben tener acceso a los datos compartidos al crear una base de datos a partir del recurso compartido de datos. Según las opciones que elija al crear la base de datos, puede controlar el acceso a la misma de la siguiente manera. Para obtener más información acerca de la creación de una base de datos a partir de un recurso compartido de datos, consulte CREATE DATABASE.

Para obtener más información sobre cómo configurar un recurso compartido de datos y leer los datos de un consumidor, consulte Uso compartido del acceso de lectura a los datos en una cuenta de AWS.

Crear la base de datos sin la cláusula WITH PERMISSIONS

El administrador puede controlar el acceso en el nivel de la base de datos o del esquema. Para controlar el acceso en el nivel de esquema, el administrador debe crear un esquema externo a partir de la base de datos de Amazon Redshift creada desde el recurso compartido de datos.

En el siguiente ejemplo, se otorgan permisos para acceder a una tabla compartida en el nivel de la base de datos y del esquema.

```
GRANT USAGE ON DATABASE sales_db TO Bob;

CREATE EXTERNAL SCHEMA sales_schema FROM REDSHIFT DATABASE sales_db SCHEMA 'public';

GRANT USAGE ON SCHEMA sales_schema TO ROLE Analyst_role;
```

Para restringir aún más el acceso, puede crear vistas encima de los objetos compartidos para así exponer solo los datos necesarios. Luego, puede utilizar estas vistas para brindar acceso a los usuarios y los roles.

Una vez que a los usuarios se les conceda acceso a la base de datos o esquema, tendrán acceso a todos los objetos compartidos de esa base de datos o esquema.

Crear la base de datos con la cláusula WITH PERMISSIONS

Tras conceder los derechos de uso en la base de datos o el esquema, un administrador puede controlar aún más el acceso mediante el mismo proceso de concesión de permisos que utilizaría en una base de datos o esquema local. Sin permisos de objetos individuales, los usuarios no pueden acceder a ningún objeto de la base de datos o el esquema de datos compartidos, incluso después de que les conceda el permiso USAGE.

En el siguiente ejemplo, se otorgan permisos para acceder a una tabla compartida en el nivel de la base de datos.

```
GRANT USAGE ON DATABASE sales_db TO Bob;
GRANT USAGE FOR SCHEMAS IN DATABASE sales_db TO Bob;
GRANT SELECT ON sales_db.public.tickit_sales_redshift TO Bob;
```

Una vez que se les haya concedido el acceso a la base de datos o al esquema, los usuarios seguirán necesitando los permisos correspondientes para cualquier objeto de la base de datos o el esquema al que desee que accedan.

Uso compartido detallado mediante WITH PERMISSIONS

Puede utilizar el uso compartido pormenorizado mediante WITH PERMISSIONS para permitir que los clústeres o los grupos de trabajo sin servidor consulten el recurso compartido de datos. En este proceso se supone que el recurso compartido de datos proviene de otro clúster o espacio de nombres de Amazon Redshift sin servidor de su cuenta, o bien proviene de otra cuenta y se ha asociado al espacio de nombres que está utilizando.

1. El administrador de la base de datos del consumidor puede crear una base de datos a partir del recurso compartido de datos.

   ```
   CREATE DATABASE my_ds_db [WITH PERMISSIONS] FROM DATASHARE my_datashare OF
    NAMESPACE 'abc123def';
   ```

   Si crea una base de datos WITH PERMISSIONS, puede conceder permisos pormenorizados sobre los objetos del recurso compartido de datos a diferentes usuarios y roles. Sin esto, a todos los usuarios y roles a los que se les conceda el permiso USAGE en la base de datos de recurso compartido de datos se les conceden todos los permisos en todos los objetos de la base de datos de recurso compartido de datos.

2. A continuación, se muestra cómo conceder permisos a un usuario o rol de base de datos de Redshift. Debe estar conectado a una base de datos local para ejecutar estas instrucciones. No puede ejecutar estas instrucciones si ejecuta un comando USE en la base de datos de recurso compartido de datos antes de ejecutar las instrucciones de concesión.

```
GRANT USAGE ON DATABASE my_ds_db TO ROLE data_eng;
GRANT CREATE, USAGE ON SCHEMA my_ds_db.my_shared_schema TO ROLE data_eng;
GRANT ALL ON ALL TABLES IN SCHEMA my_ds_db.my_shared_schema TO ROLE data_eng;

GRANT USAGE ON DATABASE my_ds_db TO bi_user;
GRANT USAGE ON SCHEMA my_ds_db.my_shared_schema TO bi_user;
GRANT SELECT ON my_ds_db.my_shared_schema.table1 TO bi_user;
```

## Eliminación de la asociación de un recurso compartido de datos de los consumidores de datos en una Cuenta de AWS diferente en Amazon Redshift

Con Amazon Redshift, puede eliminar la asociación de los recursos compartidos de datos que comparten otras Cuentas de AWS. Los recursos compartidos de datos son objetos de bases de datos que se pueden compartir y que encapsulan datos de una o más bases de datos de Redshift. En las siguientes secciones se muestra el proceso de eliminación de la asociación de los recursos compartidos de datos en su entorno de Redshift.

### Console

Como administrador de consumidores, puede eliminar la asociación de los recursos compartidos de datos de los consumidores de datos en la consola.

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. En el menú de navegación, elija Datashares (Recursos compartidos de datos). Se abrirá la página con la lista de datashares.
3. Elija From other accounts (De otras cuentas).
4. En la sección Datashares from other accounts (Datashares de otras cuentas), elija el datashare cuya asociación a los consumidores de datos desea eliminar.
5. En la sección Data consumers (Consumidores en datos), elija uno o más consumidores de datos para los cuales desea eliminar la asociación. A continuación, elija Remove association (Eliminar asociación).

6. Cuando aparezca la página de eliminación de la asociación, elija Remove association (Eliminar asociación).

Después de eliminar la asociación, los consumidores de datos perderán el acceso al datashare. Puede cambiar la asociación de consumidores de datos en cualquier momento.

SQL

> **Note**
>
> Los pasos de esta sección se llevan a cabo después de que el administrador de productores autorice acciones específicas en los objetos de base de datos compartidos y, si el recurso compartido de datos se comparte con otra cuenta, el administrador de seguridad de productores autorice el acceso.

El administrador de seguridad de consumidores puede eliminar la asociación del recurso compartido de datos con el siguiente comando:

```
disassociate-data-share-consumer
--data-share-arn <value>
```

Para obtener más información acerca del comando, consulte disassociate-data-share-consumer.

Rechazo de un recurso compartido de datos de otra Cuenta de AWS en Amazon Redshift

Con Amazon Redshift, puede rechazar los recursos compartidos de datos que comparten otras cuentas de AWS, lo que permite compartir datos de forma segura y sin problemas a través de los límites de la organización. Los recursos compartidos de datos son objetos de bases de datos que se pueden compartir y que encapsulan datos de una o más bases de datos de Amazon Redshift.

Como administrador de consumidores, puede rechazar cualquier recurso compartido de datos cuyo estado sea disponible o activo. Después de rechazar un recurso compartido de datos, los usuarios del clúster consumidor perderán el acceso a dicho recurso. Amazon Redshift no devuelve el recurso compartido de datos rechazado si se llama a la operación de la API `DescribeDataSharesForConsumer`. Si el administrador de productores ejecuta la operación de la API `DescribeDataSharesForProducer`, verá que se rechazó el recurso compartido de datos. Una vez que se rechaza un recurso compartido de datos, el administrador de productores puede

volver a autorizar el recurso compartido de datos a un clúster de consumidor y el administrador del clúster de consumidores puede decidir asociar su cuenta de AWS al recurso compartido de datos o rechazarlo.

Si su cuenta de AWS tiene una asociación a un recurso compartido de datos y una asociación pendiente a un recurso compartido de datos administrado por Lake Formation, al rechazar la asociación del recurso compartido de datos que administra Lake Formation también se rechaza el recurso compartido de datos original. Para rechazar una asociación específica, el administrador de productores puede eliminar la autorización de un recurso compartido de datos especificado. Esta acción no afecta a otros recursos compartidos de datos.

Para rechazar un recurso compartido de datos, utilice la consola de AWS, la operación de la API `RejectDataShare` o `reject-datashare` en la AWS CLI.

Console

Para rechazar un recurso compartido de datos utilizando la consola de AWS, realice los siguientes pasos.

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. En el menú de navegación, elija Recursos compartidos de datos.
3. Elija From other accounts (De otras cuentas).
4. En la sección Datashares from other accounts (Datashares de otras cuentas), elija el datashare que desee rechazar. Cuando aparezca la página Decline datashare (Rechazar recurso compartido de datos), elija Decline (Rechazar).

Después de rechazar los datashares, no puede revertir el cambio. Amazon Redshift elimina los recursos compartidos de datos de la lista. Para volver a ver el recurso compartido de datos, el administrador del productor debe volver a autorizarlo.

CLI

Para rechazar un recurso compartido de datos, el administrador utiliza el siguiente comando:

```
reject-data-share
--data-share-arn <value>
```

Para obtener más información acerca del comando, consulte reject-data-share.

# Introducción al uso compartido de datos con AWS CloudFormation en Amazon Redshift

Puede automatizar la configuración del uso compartido de datos mediante el uso de una pila de AWS CloudFormation, que aprovisiona los recursos de AWS. La pila de CloudFormation configura el uso compartido de datos entre dos clústeres de Amazon Redshift en la misma cuenta de AWS. De este modo, puede iniciar el uso compartido de datos sin ejecutar instrucciones SQL para aprovisionar los recursos.

La pila crea un recurso compartido de datos en el clúster designado. El recurso compartido de datos incluye una tabla y datos de solo lectura de muestra. Otro clúster de Amazon Redshift puede leer estos datos.

Si desea comenzar a compartir datos en una cuenta de AWS mediante la ejecución de instrucciones SQL para configurar un recurso compartido de datos y conceder permisos sin utilizar CloudFormation, consulte Uso compartido de acceso de lectura a los datos en una Cuenta de AWS.

Antes de ejecutar la pila de CloudFormation para compartir datos, debe iniciar sesión con un usuario que tenga permiso para crear un rol de IAM y una función de Lambda. También necesitas dos clústeres de Amazon Redshift en la misma cuenta. Uno, el productor, se utiliza para compartir los datos de muestra, y el otro, el consumidor, para leerlos. El requisito principal para estos clústeres es que cada uno utilice nodos RA3. Para informarse de los requisitos adicionales, consulte Consideraciones sobre el uso compartido de datos en Amazon Redshift.

Para obtener más información acerca de cómo empezar a configurar un clúster de Amazon Redshift, consulte Introducción a los almacenamientos de datos aprovisionados de Amazon Redshift. Para obtener más información acerca de la automatización de la configuración con CloudFormation, consulte ¿Qué es AWS CloudFormation?

> **Important**
>
> Antes del lanzamiento de la pila de CloudFormation, asegúrese de que tiene dos clústeres de Amazon Redshift en la misma cuenta y que los clústeres usan nodos RA3. Asegúrese de que cada clúster tenga una base de datos y un superusuario. Para obtener más información, consulte CREATE DATABASE y superuser.

Para lanzar la pila de CloudFormation para el uso compartido de datos de Amazon Redshift, haga lo siguiente:

1. Haga clic en Launch CFN stack (Lanzar la pila de CFN), lo que lo llevará al servicio de CloudFormation en la AWS Management Console.

   Si se le solicita, inicie sesión.

   El proceso de creación de la pila se inicia si hace referencia a un archivo de plantilla de CloudFormation que se almacena en Amazon S3. Una plantilla de CloudFormation es un archivo de texto en formato JSON que establece los recursos de AWS que componen una pila. Para obtener más información acerca de las plantillas de CloudFormation, consulte Más información sobre los aspectos básicos de las plantillas.
2. Elija Next (Siguiente) para ingresar los detalles de la pila.
3. En Parameters (Parámetros), para cada clúster, ingrese lo siguiente:

   - El nombre del clúster de Amazon Redshift; por ejemplo, **`ra3-consumer-cluster`**
   - El nombre de la base de datos, por ejemplo **`dev`**
   - El nombre del usuario de la base de datos; por ejemplo, **`consumeruser`**

   Recomendamos utilizar clústeres de prueba porque la pila crea varios objetos de base de datos.

   Elija Next (Siguiente).
4. Aparecerán las opciones de la pila.

   Elija Next (Siguiente) para aceptar la configuración predeterminada.
5. En Capabilities (Capacidades), elija I acknowledge that AWS CloudFormation might create IAM resources (Confirmo que AWS CloudFormation puede crear recursos de IAM).
6. Seleccione Creación de pila.

CloudFormation tarda unos 10 minutos en crear la pila de Amazon Redshift con la plantilla y se crea un recurso compartido de datos denominado `myproducer_share`. La pila crea el recurso compartido de datos en la base de datos especificada en los detalles de la pila. Solo los objetos de esa base de datos pueden compartirse.

Si se produce un error durante la creación de la pila, haga lo que se indica a continuación:

- Asegúrese de haber ingresado el nombre del clúster, el nombre de la base de datos y el nombre del usuario de la base de datos correctos para cada clúster de Redshift.
- Asegúrese de que el clúster tenga nodos RA3.
- Asegúrese de haber iniciado sesión con un usuario que tenga permiso para crear un rol de IAM y una función de Lambda. Para obtener más información acerca de cómo crear roles de IAM, consulte Creación de roles de IAM. Para obtener más información sobre las políticas de creación de la función Λ, consulte Desarrollo de funciones.

## Consulta del recurso compartido de datos creado

Para utilizar el siguiente procedimiento, asegúrese de tener los permisos necesarios para ejecutar consultas en cada clúster descrito.

Para consultar su recurso compartido de datos:

1. Conéctese al clúster productor de la base de datos que ingresó cuando creó la pila de CloudFormation, utilizando una herramienta de cliente como el Editor de consultas de Amazon Redshift versión 2.
2. Consulte los recursos compartidos de datos.

```
SHOW DATASHARES;

+------------------+-------------+-----------------+-------------------
+------------+------------+--------------------+-----------+-----------------
+-------------------------------------+
|    share_name    | share_owner | source_database | consumer_database | share_type
 | createdate | is_publicaccessible | share_acl | producer_account |
 producer_namespace          |
+------------------+-------------+-----------------+-------------------
+------------+------------+--------------------+-----------+-----------------
+-------------------------------------+
| myproducer_share | 100         | sample_data_dev | myconsumer_db     | INBOUND
  | NULL       | true                | NULL      |  producer-acct   |        your-
producer-namespace       |
+------------------+-------------+-----------------+-------------------
+------------+------------+--------------------+-----------+-----------------
+-------------------------------------+
```

El comando anterior devuelve el nombre del recurso compartido de datos que creó la pila, denominado `myproducer_share`. También devuelve el nombre de la base de datos asociada al recurso compartido de datos, `myconsumer_db`.

Copie el identificador de espacio de nombres productor para utilizarlo en un paso posterior.

3. Describa los objetos del recurso compartido de datos.

```
DESC DATASHARE myproducer_share;

+------------------+--------------------------------------+------------
+------------------+-------------+-----------------------------------
+-------------+
| producer_account |          producer_namespace          | share_type |
 share_name    | object_type |             object_name             | include_new |
+------------------+--------------------------------------+------------
+------------------+-------------+-----------------------------------
+-------------+
|   producer-acct  |        your-producer-namespace       | OUTBOUND   |
 myproducer_share | schema      | myproducer_schema                   | true
 |
|   producer-acct  |        your-producer-namespace       | OUTBOUND   |
 myproducer_share | table       | myproducer_schema.tickit_sales      | NULL
 |
|   producer-acct  |        your-producer-namespace       | OUTBOUND   |
 myproducer_share | view        | myproducer_schema.ticket_sales_view | NULL
 |
+------------------+--------------------------------------+------------
+------------------+-------------+-----------------------------------
+-------------+
```

Cuando describe el recurso compartido de datos, devuelve las propiedades de las tablas y las vistas. La pila agrega tablas y vistas con datos de muestra a la base de datos productora; por ejemplo, `tickit_sales` y `tickit_sales_view`. Para obtener más información acerca de la base de datos de ejemplo TICKIT, consulte Base de datos de muestra.

No es necesario delegar permisos en el recurso compartido de datos para ejecutar las consultas. La pila otorga los permisos necesarios.

4. Conéctese al clúster consumidor con la herramienta del cliente. Describa el recurso compartido de datos y especifique el espacio de nombres del productor.

```
DESC DATASHARE myproducer_share OF NAMESPACE '<namespace id>'; --specify the unique
 identifier for the producer namespace

+-----------------+--------------------------------------+-----------
+-----------------+-------------+-----------------------------------
+-------------+
| producer_account |          producer_namespace          | share_type |
 share_name    | object_type |             object_name             | include_new |
+-----------------+--------------------------------------+-----------
+-----------------+-------------+-----------------------------------
+-------------+
|   producer-acct  |       your-producer-namespace       | INBOUND    |
 myproducer_share | schema      | myproducer_schema                   | NULL
 |
|   producer-acct  |       your-producer-namespace       | INBOUND    |
 myproducer_share | table       | myproducer_schema.tickit_sales      | NULL
 |
|   producer-acct  |       your-producer-namespace       | INBOUND    |
 myproducer_share | view        | myproducer_schema.ticket_sales_view | NULL
 |
+-----------------+--------------------------------------+-----------
+-----------------+-------------+-----------------------------------
+-------------+
```

5. Para consultar las tablas del recurso compartido de datos, puede especificar la base de datos y el esquema del recurso compartido de datos. Para obtener más información, consulte Ejemplos de consultas entre bases de datos. Las siguientes consultas muestran datos de ventas y vendedores de la tabla VENTAS de la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

```
SELECT * FROM myconsumer_db.myproducer_schema.tickit_sales_view;

+---------+--------+----------+---------+---------+--------+---------+-----------
+------------+---------------------+
| salesid | listid | sellerid | buyerid | eventid | dateid | qtysold | pricepaid |
 commission |      saletime       |
+---------+--------+----------+---------+---------+--------+---------+-----------
+------------+---------------------+
|       1 |      1 |    36861 |   21191 |    7872 |   1875 |       4 |       728 |
     109.2 | 2008-02-18 02:36:48 |
|       2 |      4 |     8117 |   11498 |    4337 |   1983 |       2 |        76 |
      11.4 | 2008-06-06 05:00:16 |
```

```
|       3 |      5 |     1616 |   17433 |    8647 |   1983 |       2 |       350 |
      52.5 | 2008-06-06 08:26:17 |
|       4 |      5 |     1616 |   19715 |    8647 |   1986 |       1 |       175 |
     26.25 | 2008-06-09 08:38:52 |
|       5 |      6 |    47402 |   14115 |    8240 |   2069 |       2 |       154 |
      23.1 | 2008-08-31 09:17:02 |
+---------+--------+----------+---------+---------+--------+---------+-----------
+------------+---------------------+
```

> **Note**
>
> La consulta se ejecuta en la vista del esquema compartido. No puede conectarse de manera directa a las bases de datos que se hayan creado a partir de los recursos compartidos de datos. Son de solo lectura.

6. Para ejecutar una consulta que incluya agregaciones, utilice el siguiente ejemplo.

```
SELECT * FROM myconsumer_db.myproducer_schema.tickit_sales ORDER BY 1,2 LIMIT 5;

+---------+--------+----------+---------+---------+--------+---------+-----------
+------------+---------------------+
| salesid | listid | sellerid | buyerid | eventid | dateid | qtysold | pricepaid |
 commission |      saletime       |
+---------+--------+----------+---------+---------+--------+---------+-----------
+------------+---------------------+
|       1 |      1 |    36861 |   21191 |    7872 |   1875 |       4 |       728 |
     109.2 | 2008-02-18 02:36:48 |
|       2 |      4 |     8117 |   11498 |    4337 |   1983 |       2 |        76 |
      11.4 | 2008-06-06 05:00:16 |
|       3 |      5 |     1616 |   17433 |    8647 |   1983 |       2 |       350 |
      52.5 | 2008-06-06 08:26:17 |
|       4 |      5 |     1616 |   19715 |    8647 |   1986 |       1 |       175 |
     26.25 | 2008-06-09 08:38:52 |
|       5 |      6 |    47402 |   14115 |    8240 |   2069 |       2 |       154 |
      23.1 | 2008-08-31 09:17:02 |
+---------+--------+----------+---------+---------+--------+---------+-----------
+------------+---------------------+
```

La consulta devuelve datos de las ventas y los vendedores de los datos de ejemplo de TICKIT.

Para obtener más ejemplos de consultas del recurso compartido de datos, consulte Uso compartido de acceso de lectura a los datos en una Cuenta de AWS.

# Tipos de recursos compartidos de datos en Amazon Redshift

El datashare (recurso para compartir datos) es la unidad que se utiliza para compartir datos en Amazon Redshift. Utilice los recursos compartidos de datos para compartir datos en la misma Cuenta de AWS o en diferentes Cuentas de AWS. Además, puede compartir datos con fines de lectura entre distintos clústeres de Amazon Redshift.

Cada recurso compartido de datos está asociado a una base de datos específica del clúster de Amazon Redshift.

Un administrador de productores también puede crear recursos compartidos de datos para compartir los datos con otros clústeres, lo que se denomina recursos compartidos salientes. El administrador de consumidores puede recibir recursos compartidos de datos de otros clústeres, lo que se denomina recursos compartidos entrantes. Para más información sobre los productores y los consumidores, consulte Productores y consumidores de recurso compartido de datos.

Los objetos de los recursos compartidos de datos son objetos de bases de datos específicas de un clúster que los administradores de productores pueden añadir a los recursos compartidos de datos para compartirlos con los consumidores de datos. Los consumidores de datos solo pueden leer los objetos de los datashares. Algunos ejemplos de los objetos de los datashares son las tablas, las vistas y las funciones definidas por el usuario. Puede agregar objetos a los datashares mientras crea o modifica los datashares en cualquier momento.

El uso compartido de datos sigue funcionando cuando los clústeres cambian el tamaño o cuando el clúster de productores está en pausa.

Existen diferentes tipos de recursos compartidos de datos: recursos compartidos de datos estándar, recursos compartidos de datos AWS Data Exchange y recursos compartidos de datos administrados por AWS Lake Formation. En las siguientes páginas se ofrece una descripción general de cada uno de ellos.

## Recursos compartidos de datos estándar

Con recursos compartidos de datos estándar, puede compartir datos entre clústeres aprovisionados, grupos de trabajo sin servidor, zonas de disponibilidad, Cuentas de AWS y Regiones de AWS. Puede

compartir entre tipos de clústeres, así como entre clústeres aprovisionados y Amazon Redshift sin servidor.

Para compartir datos, tenga en cuenta el siguiente clúster aprovisionado, el espacio de nombres sin servidor e identificadores de Cuenta de AWS:

- Los espacios de nombres aprovisionados son los identificadores de los clústeres aprovisionados de Amazon Redshift. El identificador global único (GUID) de un espacio de nombres se crea de forma automática durante la creación del clúster aprovisionado y se adjunta al clúster. El nombre de recurso de Amazon (ARN) de un espacio de nombres está en el formato arn:{partition}:redshift:{region}:{account-id}:namespace:{namespace-guid}. Puede ver el espacio de nombres de un clúster aprovisionado en la página de detalles del clúster de la consola de Amazon Redshift.

  En el flujo de trabajo del uso compartido de datos, el valor GUID del espacio de nombres y el ARN del espacio de nombres se utilizan para compartir datos con los clústeres de la Cuenta de AWS. También puede encontrar el espacio de nombres del clúster actual con la función `current_namespace`.
- Los espacios de nombres sin servidor son los identificadores de Amazon Redshift sin servidor. El identificador global único (GUID) de un espacio de nombres se crea de forma automática durante la creación de Amazon Redshift sin servidor y también se adjunta a la instancia. El ARN de un espacio de nombres sin servidor está en el formato arn:{partition}:redshift-serverless:{region}:{account-id}:namespace/{namespace-guid}.
- Las Cuentas de AWS pueden ser consumidoras de los recursos compartidos de datos y cada una de ellas se representa con un ID de Cuenta de AWS de 12 dígitos.

Para los recursos compartidos de datos estándares, tenga en cuenta lo siguiente:

- Cuando se elimina un clúster productor, Amazon Redshift elimina los datashares creados por el clúster productor. Cuando se realiza una copia de seguridad de un clúster productor y se restaura, los datashares creados se conservan en el clúster restaurado. No obstante, los permisos para los datashares que se hayan otorgado a otros clústeres ya no son válidos en el clúster restaurado. Vuelva a otorgar permisos de uso de los datashares a los clústeres consumidores deseados. La base de datos consumidora del clúster consumidor apunta al datashare del clúster original del cual se tomó la instantánea. Para consultar los datos compartidos del clúster restaurado, el administrador de consumidores crea una base de datos diferente. O bien, el administrador puede eliminar y volver a crear una base de datos consumidora existente para utilizar el recurso compartido de datos del clúster recién restaurado.

- Cuando un clúster consumidor se elimina y se restaura a partir de una instantánea, el acceso anterior que se ha compartido con este clúster ya no será válido ni visible. Si aún se requiere el acceso a los recursos compartidos de datos en el clúster de consumidor restaurado, el administrador de productores deberá autorizar de nuevo el uso de los recursos compartidos de datos por parte del clúster de consumidor restaurado. El administrador de consumidores debe eliminar todas las bases de datos de consumidor obsoletas que se hayan creado a partir de recursos compartidos de datos inactivos. Luego, el administrador debe volver a crear la base de datos consumidora a partir del recurso compartido de datos, después de que el productor vuelva a otorgar los permisos. Dado que el GUID del espacio de nombres restaurado es diferente al del clúster original, se deben volver a conceder los permisos sobre el recurso compartido de datos cuando el clúster de consumidor o productor se restaure a partir de una copia de seguridad.

## Recursos compartidos de datos de AWS Data Exchange

Puede usar recursos compartidos de datos de AWS Data Exchange para administrar la facturación del uso compartido de datos de Amazon Redshift.

Un recurso compartido de datos de AWS Data Exchange es una unidad de licencias para compartir los datos a través de AWS Data Exchange. AWS administra toda la facturación y los pagos asociados a las suscripciones a AWS Data Exchange y el uso compartido de datos de Amazon Redshift. Los proveedores de datos aprobados pueden agregar recursos compartidos de datos de AWS Data Exchange a productos de AWS Data Exchange. Cuando los clientes se suscriben a un producto con recursos compartidos de datos de AWS Data Exchange, obtienen acceso a los recursos compartidos de datos del producto.

AWS Data Exchange for Amazon Redshift le facilita el acceso con licencias a sus datos de Amazon Redshift a través de AWS Data Exchange. Cuando un cliente se suscribe a un producto con recursos compartidos de datos de AWS Data Exchange, AWS Data Exchange agrega de forma automática al cliente como consumidor de datos en todos recursos compartidos de datos de AWS Data Exchange incluidos con el producto. Las facturas se generan de forma automática, y los pagos se cobran de forma centralizada y se abonan de forma automática a través de AWS Marketplace Entitlement Service.

Los proveedores pueden obtener licencias de datos en Amazon Redshift a un nivel pormenorizado, como en los esquemas, las tablas, las vistas y las funciones definidas por el usuario. Puede utilizar el mismo recurso compartido de datos de AWS Data Exchange en varios productos de AWS Data Exchange. Todos los objetos que se agreguen al recurso compartido de datos de AWS Data Exchange están disponibles para los consumidores. Los productores pueden ver todos los recursos

compartidos de datos de AWS Data Exchange administrados por AWS Data Exchange en su nombre mediante las operaciones de la API de Amazon Redshift, los comandos SQL y la consola de Amazon Redshift. Los clientes que se suscriben a los recursos compartidos de datos de AWS Data Exchange de un producto tienen acceso de solo lectura a los objetos de los recursos compartidos de datos.

Los clientes que deseen utilizar datos de productores de terceros pueden examinar el catálogo de AWS Data Exchange para descubrir los conjuntos de datos en Amazon Redshift y suscribirse a ellos. Una vez que su suscripción a AWS Data Exchange esté activa, pueden crear una base de datos a partir del recurso compartido de datos de su clúster y consultar los datos en Amazon Redshift.

## Funcionamiento de los recursos compartidos de datos de AWS Data Exchange

### Administración de los recursos compartidos de datos de AWS Data Exchange como administrador de un productor

Si es un productor de datos (también conocido como proveedor en AWS Data Exchange), puede crear recursos compartidos de datos de AWS Data Exchange que se conecten a sus bases de datos de Amazon Redshift. Para agregar recursos compartidos de datos de AWS Data Exchange a productos en AWS Data Exchange, debe ser un proveedor registrado de AWS Data Exchange.

Para obtener más información sobre cómo comenzar a utilizar los recursos compartidos de datos de AWS Data Exchange, consulte Uso compartido de datos de Amazon Redshift con licencia en AWS Data Exchange.

### Uso de recursos compartidos de datos de AWS Data Exchange como consumidor con una suscripción a AWS Data Exchange activa

Si es un consumidor con una suscripción a AWS Data Exchange activa (también conocido como suscriptor en AWS Data Exchange), puede examinar el catálogo de AWS Data Exchange en la consola de AWS Data Exchange para descubrir productos que contengan recursos compartidos de datos de AWS Data Exchange.

Después de suscribirse a un producto que contenga recursos compartidos de datos de AWS Data Exchange, cree una base de datos a partir del recurso compartido de datos dentro del clúster. A continuación, puede consultar los datos directamente en Amazon Redshift sin extraer, transformar y cargar los datos.

Para obtener más información sobre cómo comenzar a utilizar los recursos compartidos de datos de AWS Data Exchange, consulte Uso compartido de datos de Amazon Redshift con licencia en AWS Data Exchange.

Para los recursos compartidos de datos de AWS Data Exchange, tenga en cuenta lo siguiente:

- Cuando se elimina un clúster productor, Amazon Redshift elimina los datashares creados por el clúster productor. Cuando se realiza una copia de seguridad de un clúster productor y se restaura, los datashares creados se conservan en el clúster restaurado. Para que los suscriptores de los datos puedan seguir accediendo a ellos, cree de nuevo los recursos compartidos de datos de AWS Data Exchange y publíquelos en los conjuntos de datos del producto. La base de datos consumidora del clúster consumidor apunta al datashare del clúster original del cual se tomó la instantánea. Para consultar los datos compartidos del clúster restaurado, el administrador de consumidores crea una base de datos diferente o elimina y vuelve a crear una base de datos del consumidor existente para utilizar el recurso compartido de datos de AWS Data Exchange que se acaba de crear del clúster recién restaurado.
- Cuando se elimina un clúster de consumidor y se restaura a partir de una instantánea, el acceso anterior compartido a este clúster sigue siendo válido y visible. El administrador de consumidores debe eliminar todas las bases de datos de consumidor obsoletas que se hayan creado a partir de recursos compartidos de datos inactivos y volver a crear la base de datos de consumidor a partir del recurso compartido de datos después de que el productor vuelva a conceder los permisos. Dado que el GUID del espacio de nombres restaurado es diferente al del clúster original, se deben volver a conceder los permisos del recurso compartido de datos cuando el clúster de productor se restaure a partir de una copia de seguridad.
- Le recomendamos que no elimine su clúster si tiene algún recurso compartido de datos de AWS Data Exchange. Realizar este tipo de alteración puede infringir los términos del producto de datos en AWS Data Exchange.

## Productores y consumidores de recurso compartido de datos

Los productores de datos (también conocidos como productores de uso compartido de datos o productores de recursos compartidos de datos) son clústeres cuyos datos se desea compartir. Los administradores de productores y los propietarios de bases de datos pueden crear recursos compartidos de datos utilizando el comando CREATE DATASHARE. Puede agregar distintos objetos, como esquemas, tablas, vistas y funciones definidas por el usuario (UDF) de SQL, desde una base de datos desde la cual desea que el clúster productor comparta datos con los clústeres consumidores.

Los productores de datos (también conocidos como proveedores en AWS Data Exchange) de los recursos compartidos de datos de AWS Data Exchange pueden obtener licencias de datos a través

de AWS Data Exchange. Los proveedores aprobados pueden agregar recursos compartidos de datos de AWS Data Exchange a productos de AWS Data Exchange.

Cuando un cliente se suscribe a un producto con recursos compartidos de datos de AWS Data Exchange, AWS Data Exchange agrega de forma automática al cliente como consumidor de datos en todos los recursos compartidos de datos de AWS Data Exchange incluidos con el producto. AWS Data Exchange también elimina a todos los clientes de los recursos compartidos de datos de AWS Data Exchange cuando finaliza su suscripción. Además, AWS Data Exchange administra de forma automática la facturación, el cobro de pagos y la distribución de pagos para los productos pagos con los recursos compartidos de datos de AWS Data Exchange. Para obtener más información, consulte Recursos compartidos de datos de AWS Data Exchange. Para registrarse como proveedor de datos de AWS Data Exchange, consulte Primeros pasos como proveedor.

Los consumidores de datos (también conocidos como consumidores del uso compartido de datos o consumidores de datashares) son clústeres que reciben datashares de los clústeres productores.

Los clústeres de Amazon Redshift que comparten datos pueden estar en la misma Cuentas de AWS o en una diferente, o bien en distintas Regiones de AWS, de modo que se pueden compartir datos entre organizaciones y colaborar con terceros. Los administradores de consumidores reciben los recursos compartidos de datos para los que se les concede el uso y revisan el contenido de cada uno de ellos. Para consumir los datos compartidos, el administrador de consumidores crea una base de datos de Amazon Redshift a partir del recurso compartido de datos. A continuación, el administrador asigna permisos para la base de datos a los usuarios y los roles del clúster del consumidor. Una vez otorgados los permisos, los usuarios y los roles pueden enumerar los objetos compartidos como parte de las consultas de metadatos estándar, junto con los datos locales del clúster del consumidor. Pueden comenzar a efectuar consultas de inmediato.

Si es un consumidor con una suscripción activa a AWS Data Exchange (también conocidos como suscriptores en AWS Data Exchange), puede encontrar y consultar datos pormenorizados y actualizados en Amazon Redshift, además de suscribirse a ellos, sin necesidad de extraer, transformar ni cargar los datos. Para obtener más información, consulte Recursos compartidos de datos de AWS Data Exchange.

## Recursos compartidos de datos administrados por AWS Lake Formation

Con Amazon Redshift, puede acceder a los datos en directo entre cuentas de AWS y clústeres de Amazon Redshift y compartirlos mediante recursos compartidos de datos administrados por AWS Lake Formation. Los recursos compartidos de datos de AWS Lake Formation permiten a los

proveedores de datos compartir de forma segura los datos en directo del lago de datos de Amazon S3 con cualquier consumidor, incluidas otras cuentas de AWS y clústeres de Amazon Redshift.

Con AWS Lake Formation, puede definir y aplicar de forma centralizada los permisos de acceso a nivel de base de datos, tabla, columna y fila de los recursos compartidos de datos de Amazon Redshift y restringir el acceso de los usuarios a los objetos dentro de un recurso compartido de datos. Al compartir datos a través de Lake Formation, puede definir los permisos en Lake Formation y aplicarlos a cualquier recurso compartido de datos y sus objetos. Por ejemplo, si tiene una tabla que contiene información de los empleados, puede usar los filtros de nivel de columna de Lake Formation para evitar que los empleados que no trabajan en el departamento de Recursos Humanos vean información de identificación personal (PII), por ejemplo, un número de la seguridad social. Para obtener más información sobre los filtros de datos, consulte Filtrado de datos y seguridad en el nivel de celdas en Lake Formation en la Guía para desarrolladores de AWS Lake Formation.

También puede utilizar etiquetas en Lake Formation para configurar los permisos en los recursos de Lake Formation. Para obtener más información, consulte Control de acceso basado en etiquetas de Lake Formation.

Actualmente, Amazon Redshift admite el uso de recursos compartidos de datos a través de Lake Formation cuando se comparten en la misma cuenta o entre varias cuentas. No se admite el uso compartido entre varias regiones.

A continuación, se brinda información general sobre cómo utilizar Lake Formation para controlar los permisos de los recursos compartidos de datos:

1. En Amazon Redshift, el administrador del clúster o grupo de trabajo del productor crea un recurso compartido de datos en el clúster o grupo de trabajo del productor y concede el uso a una cuenta de Lake Formation.
2. El administrador del clúster o grupo de trabajo del productor autoriza a la cuenta de Lake Formation a acceder al recurso compartido de datos.
3. El administrador de Lake Formation descubre y registra los recursos compartidos de datos. También debe descubrir los ARN de AWS Glue a los que tiene acceso y asociar los datos compartidos a un ARN de AWS Glue Data Catalog. Si utiliza la AWS CLI, puede descubrir y aceptar recursos compartidos de datos con las operaciones `describe-data-shares` y `associate-data-share-consumer` de la CLI de Redshift. Para registrar un recurso compartido de datos, utilice la operación `register-resource` de la CLI de Lake Formation.
4. El administrador de Lake Formation crea una base de datos federada en AWS Glue Data Catalog y configura los permisos de Lake Formation para controlar el acceso de los usuarios a

los objetos del recurso compartido de datos. Para obtener más información sobre las bases de datos federadas en AWS Glue, consulte Administración de permisos para datos en un recurso compartido de datos de Amazon Redshift.

5. El administrador de Lake Formation descubre las bases de datos de AWS Glue a las que tiene acceso y asocia el recurso compartido de datos a un ARN de AWS Glue Data Catalog.
6. El administrador de Redshift descubre los ARN de la base de datos de AWS Glue a los que tiene acceso, crea una base de datos externa en el clúster consumidor de Amazon Redshift mediante un ARN de base de datos de AWS Glue y concede el uso a los usuarios de base de datos autenticados con credenciales de IAM para empezar a consultar la base de datos de Amazon Redshift.
7. Los usuarios de la base de datos pueden usar las vistas SVV_EXTERNAL_TABLES y SVV_EXTERNAL_COLUMNS para buscar todas las tablas o columnas de la base de datos de AWS Glue a las que tienen acceso y, a continuación, consultar las tablas de la base de datos de AWS Glue.
8. Cuando el administrador del clúster o grupo de trabajo de productor decide dejar de compartir los datos con el clúster de consumidor, el administrador de productores puede revocar el uso, desautorizar o eliminar el recurso compartido de datos de Redshift. Los permisos y los objetos asociados en Lake Formation no se eliminan automáticamente.

Para obtener más información sobre cómo compartir un recurso compartido de datos con AWS Lake Formation como administrador del clúster o grupo de trabajo del productor, consulte Uso de recursos compartidos de datos administrados por Lake Formation como productor. Para consumir los datos compartidos del clúster o grupo de trabajo del productor, consulte Uso de recursos compartidos de datos administrados por Lake Formation como consumidor.

# Valores de estado de recurso compartido de datos en Amazon Redshift

Con Amazon Redshift, puede compartir datos en directo de forma segura entre clústeres de Amazon Redshift sin tener que copiar ni transferir datos. Los recursos compartidos de datos para Amazon Redshift le permiten compartir los resultados de las consultas en directo, incluidas las actualizaciones de los datos de origen, con cualquier clúster de Amazon Redshift de la misma o de diferentes cuentas de AWS y regiones de AWS. En este tema se describen los posibles estados que pueden tener los recursos compartidos de datos en Amazon Redshift.

Respecto de los datashares entre cuentas, existen diferentes estados de dichos datashares que requieren sus acciones. El recurso compartido de datos puede tener un estado activo, de acción necesaria o inactivo.

A continuación, se describe cada estado de los datashares y la acción que requiere:

- Cuando un administrador de productores crea un recurso compartido de datos, el estado de ese recurso compartido de datos en el clúster de productor es Autorización pendiente. El administrador de productores puede autorizar a los consumidores de datos a acceder al recurso compartido de datos. No se necesita ninguna acción para el administrador de consumidores.
- Cuando un administrador de productores autoriza el recurso compartido de datos, el estado de dicho recurso compartido de datos pasa a ser Autorizado en el clúster productor. No se necesita ninguna acción para el administrador de productores. Cuando hay al menos una asociación a un consumidor de datos para el datashare, el estado del datashare pasa de Authorized (Autorizado) a Active (Activo).

  El estado del datashare pasa a ser Available (Action required on the Amazon Redshift console) (Disponible [Se requiere acción en la consola de Amazon Redshift]) en el clúster consumidor. El administrador de consumidores puede asociar el recurso compartido de datos con los consumidores de los datos o rechazar el recurso compartido de datos. El administrador de consumidores también puede usar el comando `describeDatashareforConsumer` de la AWS CLI para ver el estado de los recursos compartidos de datos. O bien, el administrador puede utilizar el comando `describeDatashare` de la CLI y proporcionar el nombre de recurso de Amazon (ARN) del datashare para ver su estado.
- Cuando el administrador de consumidores asocia un recurso compartido de datos con los consumidores de datos, el estado de dicho recurso compartido de datos pasa a ser Activo en el clúster de productor. Cuando hay al menos una asociación a un consumidor de datos para el datashare, el estado del datashare pasa de Authorized (Autorizado) a Active (Activo). No se requiere ninguna acción para el administrador de productores.

  El estado del datashare pasa a ser Active (Activo) en el clúster consumidor. No se requiere ninguna acción para el administrador de consumidores.
- Cuando el administrador de consumidores cancela una asociación de consumidor de un recurso compartido de datos, el estado de dicho recurso compartido de datos pasa a ser Activo o Autorizado. Pasa a ser Active (Activo) cuando existe al menos una asociación del datashare a otro consumidor de datos. Pasa a ser Authorized (Autorizado) cuando no hay ninguna asociación

de consumidores al datashare en el clúster productor. No se necesita ninguna acción para el administrador de productores.

El estado del datashare pasa a ser Action required (De acción requerida) en el clúster consumidor si se cancelan todas las asociaciones. El administrador de consumidores puede volver a asociar el recurso compartido de datos con los consumidores de datos cuando el recurso compartido de datos esté disponible para ellos.

- Cuando el administrador de consumidores rechaza un recurso compartido de datos, el estado de dicho recurso compartido de datos pasa a ser Se requiere una acción en el clúster de productor y Rechazado en el clúster de consumidor. El administrador de productores puede volver a autorizar el recurso compartido de datos. No se necesita ninguna acción para el administrador de consumidores.
- Cuando el administrador de productores cancela la autorización de un recurso compartido de datos, el estado de dicho recurso compartido de datos pasa a ser Se requiere una acción en el clúster de productor. El administrador de productores puede decidir volver a autorizar el recurso compartido de datos, si fuera necesario. No se requiere ninguna acción para el administrador de consumidores.

# Administración del acceso a las operaciones de la API de uso compartido de datos con las políticas de IAM

Para controlar el acceso a las operaciones de la API de uso compartido de datos, utilice políticas basadas en acciones de IAM. Para obtener información sobre cómo administrar las políticas de IAM, consulte Administración de las políticas de IAM en la Guía del usuario de IAM.

Para obtener información sobre los permisos necesarios para utilizar las operaciones de la API de uso compartido de datos, consulte Permisos necesarios para utilizar las operaciones de la API de uso compartido de datos en la Guía de administración de Amazon Redshift.

Para que el uso compartido de datos entre cuentas sea más seguro, puede utilizar una clave condicional `ConsumerIdentifier` para las operaciones de la API `AuthorizeDataShare` y `DeauthorizeDataShare`. De este modo, puede controlar de manera explícita qué Cuentas de AWS pueden realizar llamadas a las dos operaciones de la API.

Puede denegar la autorización del uso compartido de datos, o desautorizarlo, para cualquier consumidor que no sea su propia cuenta. Para ello, especifique el número de Cuenta de AWS en la política de IAM.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "VisualEditor0",
            "Effect": "Deny",
            "Action": [
                "redshift:AuthorizeDataShare",
                "redshift:DeauthorizeDataShare"
            ],
            "Resource": "*",
            "Condition": {
                "StringNotEquals": {
                    "redshift:ConsumerIdentifier": "555555555555"
                }
            }
        }
    ]
}
```

Puede permitir que un productor con un DatasHareARarn **`testshare2`** comparta de forma explícita con un consumidor con una Cuenta de AWS de 111122223333 en la política de IAM.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "VisualEditor0",
            "Effect": "Allow",
            "Action": [
                "redshift:AuthorizeDataShare",
                "redshift:DeauthorizeDataShare"
            ],
            "Resource": "arn:aws:redshift:us-
east-1:666666666666:datashare:af06285e-8a45-4ee9-b598-648c218c8ff1/testshare2",
            "Condition": {
                "StringEquals": {
                    "redshift:ConsumerIdentifier": "111122223333"
                }
            }
        }
    ]
```

```
}
```

# Conexión a bases de datos de consumidores de Amazon Redshift

Con una conexión directa a una base de datos de recurso compartido de datos, puede conectarse directamente a una base de datos creada a partir de un recurso compartido de datos de la misma manera que puede conectarse a cualquier otro tipo de base de datos de Amazon Redshift. Por ejemplo, puede conectarse a una base de datos creada a partir de un recurso compartido de datos mediante controladores JDBC u ODBC, el editor de consultas V2 de Amazon Redshift o cualquier otra herramienta que pueda conectarse a una base de datos de Amazon Redshift. Para obtener más información, consulte Conexión a un almacenamiento de datos de Amazon Redshift mediante herramientas de cliente de SQL.

## Acceso a datos compartidos

Cuando se conecte a una base de datos creada de un recurso compartido de datos, puede consultar los objetos compartidos con la notación de dos partes (*schema_name*.*table_name*). Si la tabla se puede encontrar en la ruta de búsqueda de la base de datos de consumidores, también puede utilizar la notación de una parte (*table_name*).

Si desea realizar consultas entre bases de datos, puede utilizar la notación de tres partes (*consumer_database_name*.*schema_name*.*table_name*). Estas consultas pueden hacer referencia a objetos compartidos de otras bases de datos de consumidores del clúster o a objetos locales de bases de datos locales. Pueden también hacer referencia a bases de datos locales y datos compartidos de otros clústeres dentro de la misma consulta.

> **Note**
>
> Una base de datos que se crea a partir de un recurso compartido de datos no tiene un catálogo local. Por lo tanto, cualquier consulta que acceda a las tablas del catálogo local, como `pg_class`, devolverá un resultado vacío.

## Acceso a los metadatos para objetos compartidos

Para ayudar a los administradores de clúster a detectar objetos compartidos en la base de datos de consumidores, Amazon Redshift proporciona un conjunto de vistas de metadatos y comandos SHOW que muestran los metadatos para estos objetos. Cuando se conecta a una base de datos de

consumidores, estas vistas y comandos de metadatos no admiten la detección de metadatos entre bases de datos. Solo devuelven los metadatos de los objetos compartidos del recurso compartido de datos que están asociados a la base de datos conectada.

Utilice SHOW SCHEMAS para ver una lista de los esquemas compartidos en el recurso compartido de datos asociado a la base de datos conectada. Para obtener más información, consulte MOSTRAR ESQUEMAS.

Utilice SHOW TABLES para ver una lista de las tablas en un esquema compartido del recurso compartido de datos asociado a la base de datos conectada. Para obtener más información, consulte SHOW TABLES.

Utilice SHOW COLUMNS para ver una lista de las columnas de una tabla compartida del recurso compartido de datos asociado a la base de datos conectada. Para obtener más información, consulte SHOW COLUMNS.

Utilice SVV_ALL_SCHEMAS para ver una lista de los esquemas compartidos en el recurso compartido de datos asociado a la base de datos conectada. Para obtener más información, consulte SVV_ALL_SCHEMAS.

Utilice SVV_ALL_TABLES para ver una lista de las tablas compartidas en el recurso compartido de datos asociado a la base de datos conectada. Para obtener más información, consulte SVV_ALL_TABLES.

Utilice SVV_ALL_COLUMNS para ver una lista de las columnas compartidas en el recurso compartido de datos asociado a la base de datos conectada. Para obtener más información, consulte SVV_ALL_COLUMNS.

## Integración del uso compartido de datos de Amazon Redshift a las herramientas de inteligencia empresarial

Para integrar el uso compartido de datos a las herramientas de inteligencia empresarial (BI), le recomendamos utilizar los controladores JDBC u ODBC de Amazon Redshift. Los controladores JDBC y ODBC de Amazon Redshift admiten la operación de la API `GetCatalogs` en los controladores. Esta operación devuelve una lista de todas las bases de datos, incluidas las creadas a partir de recursos compartidos de datos.

Los controladores también admiten las operaciones posteriores, como `GetSchemas` y GetTables, que devuelven datos de todas las bases de datos que `GetCatalogs` devuelve. Los controladores

proporcionan esta compatibilidad incluso cuando no especifica de forma explícita el catálogo en la llamada. Para obtener más información acerca de los controladores JDBC u ODBC, consulte Configuración de conexiones en la Guía de administración de Amazon Redshift.

El editor de consultas V2 de Amazon Redshift incluye bases de datos de consumidores en su interfaz de cambio de conexión. Sin embargo, la mayoría de las herramientas excluyen estas bases de datos y solo incluyen las bases de datos de clústeres locales como bases de datos conectables.

> **Note**
>
> Se ha agregado una nueva base de datos del sistema denominada `sys:internal` para el mantenimiento interno. Algunas herramientas incluyen esta base de datos del sistema como una base de datos conectable. Sin embargo, no puede conectarse a ella o ejecutar consultas con sus objetos.

# Monitoreo y auditoría en el uso compartido de datos en Amazon Redshift

Con Amazon Redshift, puede supervisar y auditar las actividades de uso compartido de datos para garantizar la conformidad y la seguridad.

Cuando se audita el uso compartido de datos, los productores pueden realizar un seguimiento de la evolución del datashare. Por ejemplo, la auditoría ayuda a realizar un seguimiento de cuándo se crean recursos compartidos de datos, se agregan o eliminan objetos y se conceden o revocan permisos a clústeres de Amazon Redshift, cuentas de AWS o regiones de AWS.

Además de la auditoría, los productores y los consumidores realizan un seguimiento del uso de los recursos compartidos de datos en diversos niveles de detalle, como los niveles de cuenta, clúster y objeto. Para obtener más información acerca del seguimiento de las vistas de uso y auditoría, consulte SVL_DATASHARE_CHANGE_LOG y SVL_DATASHARE_USAGE_PRODUCER.

Puede monitorear los recursos compartidos de datos consultando las vistas del sistema.

1. El administrador de productores que desea compartir datos crea un recurso compartido de datos de Amazon Redshift. Luego, añade los objetos de base de datos necesarios. Estos pueden ser esquemas, tablas y vistas del recurso compartido de datos, y se especifica una lista de consumidores con los que se compartirán los objetos.

Utilice las siguientes vistas del sistema para ver vistas consolidadas para realizar un seguimiento de los cambios y el uso de los recursos compartidos de datos en los clústeres de productores o consumidores:

- SYS_DATASHARE_CHANGE_LOG
- SYS_DATASHARE_USAGE_CONSUMER
- SYS_DATASHARE_USAGE_PRODUCER

Utilice las siguientes vistas de sistema para consultar los objetos del datashare y la información del consumidor de datos de los datashares salientes:

- SVV_DATASHARES
- SVV_DATASHARE_CONSUMERS
- SVV_DATASHARE_OBJECTS

2. Los administradores de consumidores examinan los recursos compartidos de datos para los que se les concede permiso de uso y revisan el contenido de cada uno de ellos visualizado los recursos compartidos de datos entrantes con SVV_DATASHARES.

   Para consumir los datos compartidos, el administrador de consumidores crea una base de datos de Amazon Redshift a partir del recurso compartido de datos. A continuación, el administrador asigna permisos a los usuarios y los roles adecuados del clúster del consumidor. Los usuarios y los roles pueden enumerar los objetos compartidos como parte de las consultas de metadatos estándar mediante la visualización de las siguientes vistas de sistema de metadatos y pueden comenzar a consultar los datos inmediatamente.

   - SVV_REDSHIFT_COLUMNS
   - SVV_REDSHIFT_DATABASES
   - SVV_REDSHIFT_FUNCTIONS
   - SVV_REDSHIFT_SCHEMAS
   - SVV_REDSHIFT_TABLES

   Para ver objetos de esquemas locales y compartidos de Amazon Redshift, además de esquemas externos, utilice las siguientes vistas de sistema de metadatos para consultarlos.

   - SVV_ALL_COLUMNS
   - SVV_ALL_SCHEMAS
   - SVV_ALL_TABLES

Cuando se conecta a una base de datos de consumidores, se desactiva la detección entre bases de datos. Las vistas del sistema de metadatos solo devuelven los metadatos para los objetos compartidos en el recurso compartido de datos asociados a la base de datos conectada.

## Integración del uso compartido de datos de Amazon Redshift con AWS CloudTrail

El uso compartido de datos se integra a AWS CloudTrail. CloudTrail es un servicio que proporciona un registro de las acciones realizadas por un usuario, un rol o un servicio de AWS en Amazon Redshift. CloudTrail captura las llamadas a la API del uso compartido de datos como eventos. Las llamadas capturadas incluyen llamadas desde la consola de AWS CloudTrail y llamadas de código a las operaciones de uso compartido de datos. Para obtener más información acerca de la integración de Amazon Redshift con AWS CloudTrail, consulte Registro con CloudTrail.

Para obtener más información sobre CloudTrail, consulte Funcionamiento de CloudTrail.

# Datos semiestructurados en Amazon Redshift

Mediante el uso de la compatibilidad con datos semiestructurados de Amazon Redshift, puede capturar y almacenar datos semiestructurados en sus almacenamientos de datos de Amazon Redshift. Mediante el uso del tipo de datos SUPER y del lenguaje PartiQL, Amazon Redshift amplía la capacidad de almacenamiento de datos para integrarse a orígenes de datos SQL y NoSQL. De esta forma, Amazon Redshift permite realizar análisis eficientes de los datos relacionales y semiestructurados almacenados, como JSON.

Amazon Redshift ofrece dos formas de compatibilidad con datos semiestructurados: el tipo de datos SUPER y Amazon Redshift Spectrum.

Utilice el tipo de datos SUPER si necesita insertar o actualizar pequeños lotes de datos JSON con baja latencia. Además, utilice SUPER cuando su consulta requiera consistencia sólida, rendimiento predecible de consultas, compatibilidad con consultas complejas y facilidad de uso con esquemas en evolución y datos sin esquemas.

Por el contrario, utilice Amazon Redshift Spectrum con un formato abierto de archivo si su consulta de datos requiere integración a otros servicios de AWS y a datos almacenados principalmente en Amazon S3 para fines de archivo.

## Casos de uso del tipo de datos SUPER

La compatibilidad con datos semiestructurados mediante el tipo de datos SUPER de Amazon Redshift proporciona rendimiento superior, flexibilidad y facilidad de uso. Los siguientes casos de uso ayudan a demostrar cómo puede usar la compatibilidad con datos semiestructurados con SUPER.

Inserción rápida y flexible de datos JSON: Amazon Redshift admite las transacciones rápidas que pueden analizar JSON y almacenarlo como un valor SUPER. Las transacciones de inserción pueden funcionar hasta cinco veces más rápido que la ejecución de las mismas inserciones en tablas que hayan fragmentado los atributos de SUPER en columnas convencionales. Por ejemplo, supongamos que el JSON entrante tiene la forma {“a”:.., “b”:.., “c”“..., ...}. Puede acelerar el rendimiento de la inserción muchas veces almacenando el JSON entrante en una tabla TJ con una única columna S de SUPER, en lugar de almacenarlo en una tabla convencional TR con columnas “a”, “b”, “c”, etc. Cuando hay cientos de atributos en el JSON, la ventaja de rendimiento del tipo de datos SUPER se vuelve sustancial.

Además, el tipo de datos SUPER no necesita un esquema normal. No es necesario realizar una introspección del JSON entrante ni limpiarlo antes de almacenarlo. Por ejemplo, supongamos que un JSON entrante tiene un atributo de cadena “c” y otros que tienen un atributo entero “c”, sin el tipo de datos SUPER. En este caso, debe separar las columnas c_string y c_int o limpiar los datos. Por el contrario, con el tipo de datos SUPER, todos los datos JSON se almacenan durante la ingesta sin perder información. Más tarde, puede usar la extensión PartiQL de SQL para analizar la información.

Consultas flexibles para la detección: después de haber almacenado los datos semiestructurados (por ejemplo, JSON) en un valor de datos SUPER, puede consultarlos sin imponer un esquema. Puede usar la escritura dinámica de PartiQL y la semántica laxa para ejecutar sus consultas y detectar los datos profundamente anidados que necesita, sin la necesidad de imponer un esquema antes de realizar la consulta.

Consultas flexibles para las operaciones de extracción, transformación y carga (ETL) en vistas materializadas convencionales: después de almacenar los datos sin esquema y semiestructurados como SUPER, puede utilizar las vistas materializadas de PartiQL para realizar una introspección de los datos y fragmentarlos en vistas materializadas.

Las vistas materializadas con los datos fragmentados son un buen ejemplo de las ventajas de rendimiento y capacidad de uso para sus casos de análisis clásicos. Cuando lleva a cabo el análisis de los datos fragmentados, la organización en columnas de las vistas materializadas de Amazon Redshift proporciona un mejor rendimiento. Además, los usuarios y las herramientas de inteligencia empresarial (BI) que requieren un esquema convencional para los datos capturados pueden utilizar vistas (ya sea materializadas o virtuales) como la presentación del esquema convencional de los datos.

Después de que las vistas materializadas de PartiQL hayan extraído los datos encontrados en JSON o SUPER en vistas materializadas convencionales organizadas en columnas, podrá consultar las vistas materializadas. Para obtener más información acerca de cómo funciona el tipo de datos SUPER con las vistas materializadas, consulte Tipo de datos SUPER y vistas materializadas.

Puede aplicar políticas de enmascaramiento dinámico de datos a los valores `scalar` de las rutas de las columnas de tipo SUPER. Para obtener más información sobre el enmascaramiento dinámico de datos, consulte Enmascaramiento de datos dinámico. Para obtener más información acerca del uso del enmascaramiento dinámico de datos con el tipo de datos SUPER, consulte Uso del enmascaramiento dinámico de datos con rutas de tipos de datos SUPER. (versión preliminar)

Para obtener más información acerca del tipo de datos SUPER, consulte Tipo SUPER.

Para obtener ejemplos de cómo utilizar el tipo de datos SUPER, consulte las subsecciones de este tema, que comienzan con Conjunto de datos de muestra SUPER.

# Conceptos para el uso del tipo de datos SUPER

A continuación, encontrará algunos conceptos acerca del tipo de datos SUPER de Amazon Redshift.

Comprensión de qué es el tipo de datos SUPER en Amazon Redshift: SUPER es un tipo de datos de Amazon Redshift que permite el almacenamiento de matrices y estructuras sin esquema que contienen escalares de Amazon Redshift y, posiblemente, matrices y estructuras anidadas. El tipo de datos SUPER puede almacenar de forma nativa diferentes formatos de datos semiestructurados, como JSON o datos procedentes de fuentes orientadas a documentos. Puede agregar una nueva columna SUPER para almacenar datos semiestructurados y escribir consultas que accedan a la columna SUPER, junto con las columnas escalares habituales. Para obtener más información acerca del tipo de datos SUPER, consulte Tipo SUPER.

Ingesta de datos JSON sin esquema como SUPER: con el tipo flexible de datos semiestructurados SUPER, Amazon Redshift puede recibir y capturar datos JSON sin esquema y pasarlos a un valor SUPER. Por ejemplo, Amazon Redshift puede capturar el valor JSON [10.5, "first"] como un valor SUPER [10.5, ‘first’], es decir, una matriz que contenga el decimal 10,5 de Amazon Redshift y el varchar “first”. Amazon Redshift puede capturar el JSON como un valor SUPER usando el comando COPY o la función JSON de análisis, como json_parse('[10.5, "first"]'). Tanto COPY como `json_parse` capturan datos JSON usando semántica de análisis estricta de manera predeterminada. También puede construir valores SUPER, incluidos matrices y estructuras, mediante los propios datos de la base de datos.

La columna SUPER no requiere modificaciones de esquema mientras se capturan las estructuras irregulares de JSON sin esquema. Por ejemplo, cuando analiza una secuencia de clics, inicialmente almacena en la columna SUPER estructuras de “clic” con atributos de “IP” y “tiempo”. Puede agregar un atributo “ID de cliente” sin cambiar el esquema para capturar dichos cambios.

El formato nativo utilizado para el tipo de datos SUPER es un formato binario que requiere menos espacio que el valor JSON en su forma textual. Esto permite una ingesta y un procesamiento de tiempo de ejecución más rápidos de los valores SUPER en la consulta.

Consulta de los datos SUPER con PartiQL: PartiQL es una extensión compatible con versiones anteriores de SQL-92 que muchos servicios de AWS utilizan actualmente. Con el uso de PartiQL, las construcciones SQL conocidas combinan sin problemas el acceso a los datos SQL tabulares clásicos

y a los datos semiestructurados de SUPER. Puede realizar la navegación de objetos y matrices y quitar el anidamiento de las matrices. PartiQL amplía el lenguaje SQL estándar para expresar y procesar declarativamente datos anidados y multivalor.

PartiQL es una extensión de SQL en la que los datos anidados y sin esquema de las columnas SUPER son ciudadanos de primera clase. PartiQL no requiere que se verifique el tipo de todas las expresiones de consulta durante el tiempo de compilación de consultas. Este enfoque permite que las expresiones de consulta que contienen el tipo de datos SUPER se escriban dinámicamente durante la ejecución de la consulta cuando se accede a los tipos reales de datos dentro de las columnas SUPER. Además, PartiQL funciona en un modo laxo, en el cual las inconsistencias de tipo no causan errores, sino que devuelven valores nulos. La combinación del procesamiento de consultas sin esquema y laxo hace que PartiQL sea ideal para las aplicaciones de extracción, carga y transferencia (ELT) en las que su consulta SQL evalúa los datos JSON que se capturan en las columnas SUPER.

Integración a Redshift Spectrum: Amazon Redshift admite varios aspectos de PartiQL al ejecutar consultas de Redshift Spectrum a través de JSON, Parquet y otros formatos que tienen datos anidados. Redshift Spectrum solo admite datos anidados que tengan esquemas. Por ejemplo, con Redshift Spectrum puede declarar que sus datos JSON tienen un atributo nested_schemaful_example en un esquema ARRAY<STRUCT<a:INTEGER, b:DECIMAL(5,2)>>. El esquema de este atributo determina que los datos siempre contengan una matriz que tenga una estructura con el entero a y el decimal b. Si los datos cambian para incluir más atributos, el tipo también cambia. Por el contrario, el tipo de datos SUPER no requiere ningún esquema. Puede almacenar matrices con elementos de estructura que tengan diferentes atributos o tipos. Además, algunos valores se pueden almacenar fuera de las matrices.

Para obtener más información acerca de las funciones que admiten el tipo de datos SUPER, consulte lo siguiente:

- Función ABS
- Función CEILING (o CEIL)
- Función FLOOR
- Función ROUND
- Función SIGN
- Función TRUNC

# Consideraciones sobre los datos SUPER

Cuando trabaje con datos SUPER, tenga presente lo siguiente:

- Utilice la versión 1.2.50 del controlador JDBC, la versión 1.4.17 del controlador ODBC o una posterior y la versión 2.0.872 del controlador Python de Amazon Redshift o una posterior.

  Para obtener más información acerca de los controladores JDBC, consulte Configuración de una conexión JDBC.

  Para obtener más información acerca de los controladores ODBC, consulte Configuración de una conexión ODBC.
- Busque los ejemplos de esquemas utilizados en los siguientes temas en Conjunto de datos de muestra SUPER.
- Todos los ejemplos de código SQL utilizados en los temas siguientes se incluyen con el mismo prefijo de S3 para descargar. Estos incluyen el lenguaje de definición de datos (DDL) y las instrucciones COPY, así como ciertas consultas modificadas por TPC-H que funcionan con SUPER.

  Para ver o descargar los archivos SQL, realice alguna de las siguientes operaciones:

  - Descargue el archivo SQL del tutorial SUPER y el archivo TPC-H.
  - Mediante la CLI de Amazon S3, ejecute el siguiente comando. Puede utilizar su propia ruta de destino.

```
aws s3 cp s3://redshift-downloads/semistructured/tutorialscripts/semistructured-
tutorial.sql /target/path
aws s3 cp s3://redshift-downloads/semistructured/tutorialscripts/
super_tpch_queries.sql /target/path
```

Para obtener más información acerca de las configuraciones de SUPER, consulte Configuraciones SUPER.

# Conjunto de datos de muestra SUPER

Puede explorar y analizar el conjunto de datos de ejemplo SUPER, que contiene datos relacionados con las ventas de productos ficticios en varias categorías, regiones y periodos de tiempo. En las

siguientes secciones se proporcionan detalles sobre el acceso, la consulta y la manipulación del conjunto de datos de ejemplo SUPER dentro del clúster de Amazon Redshift.

El esquema de la tabla y el modelo de datos utilizados para los ejemplos de ingesta y consulta se definen de la siguiente manera.

```
/*customer-orders-lineitem*/
CREATE TABLE customer_orders_lineitem
(c_custkey bigint
,c_name varchar
,c_address varchar
,c_nationkey smallint
,c_phone varchar
,c_acctbal decimal(12,2)
,c_mktsegment varchar
,c_comment varchar
,c_orders super
);

/* Datamodel of documents to be stored in c_orders Super column would be as follows*/
ARRAY < STRUCT < o_orderkey:bigint
                         ,o_orderstatus:string
                         ,o_totalprice:double
                         ,o_orderdate:string
                         ,o_orderpriority:string
                         ,o_clerk:string
                         ,o_shippriority:int
                         ,o_comment:string
                         ,o_lineitems:ARRAY < STRUCT < l_partkey:bigint
                                                      ,l_suppkey:bigint
                                                      ,l_linenumber:int
                                                      ,l_quantity:double
                                                      ,l_extendedprice:double
                                                      ,l_discount:double
                                                      ,l_tax:double
                                                      ,l_returnflag:string
                                                      ,l_linestatus:string
                                                      ,l_shipdate:string
                                                      ,l_commitdate:string
                                                      ,l_receiptdate:string
                                                      ,l_shipinstruct:string
                                                      ,l_shipmode:string
                                                      ,l_comment:string
```

```
                                                                > >
                                > >

/*part*/
CREATE TABLE part
(
 p_partkey bigint
 ,p_name varchar
 ,p_mfgr varchar
 ,p_brand varchar
 ,p_type varchar
 ,p_size int
 ,p_container varchar
 ,p_retailprice decimal(12,2)
 ,p_comment varchar
 );

/*region-nations*/
CREATE TABLE region_nations
(
 r_regionkey smallint
 ,r_name varchar
 ,r_comment varchar
 ,r_nations super
 );

 /* Datamodel of documents to be stored in r_nations Super column would be as follows*/
 ARRAY < STRUCT < n_nationkey:int,n_name:string,n_comment:string > >

/*supplier-partsupp*/
CREATE TABLE supplier_partsupp
(
 s_suppkey bigint
 ,s_name varchar
 ,s_address varchar
 ,s_nationkey smallint
 ,s_phone varchar
 ,s_acctbal double precision
 ,s_comment varchar
 ,s_partsupps super
 );

 /* Datamodel of documents to be stored in s_partsupps Super column would be as
 follows*/
```

```
ARRAY < STRUCT <
ps_partkey:bigint,ps_availqty:int,ps_supplycost:double,ps_comment:string > >
```

# Carga de datos semiestructurados en Amazon Redshift

Utilice el tipo de datos SUPER para continuar y consultar datos jerárquicos y genéricos en Amazon Redshift. Amazon Redshift presenta la función `json_parse` para analizar datos en formato JSON y convertirlos en la representación SUPER. Amazon Redshift también admite la carga de columnas SUPER mediante el uso del comando COPY. Los formatos de archivos admitidos son JSON, Avro, texto, formato de valores separados por comas (CSV), Parquet y ORC.

Para obtener información acerca de las tablas utilizadas en los ejemplos siguientes, consulte Conjunto de datos de muestra SUPER.

Para obtener más información acerca de la función `json_parse`, consulte Función JSON_PARSE.

La codificación predeterminada para el tipo de datos SUPER es ZSTD.

## Análisis de documentos JSON en columnas SUPER

Puede insertar o actualizar datos JSON en una columna SUPER mediante la función `json_parse`. La función analiza los datos en formato JSON y los convierte en el tipo de datos SUPER, que puede utilizar en las instrucciones INSERT o UPDATE.

En el siguiente ejemplo, se insertan datos JSON en una columna SUPER. Si la función `json_parse` falta en la consulta, Amazon Redshift trata el valor como una cadena simple en lugar de como una cadena con formato JSON que debe analizarse.

Si actualiza una columna de datos SUPER, Amazon Redshift requiere que el documento completo se transfiera a valores de columna. Amazon Redshift no admite las actualizaciones parciales.

```
INSERT INTO region_nations VALUES(0,
   'lar deposits. blithely final packages cajole. regular waters are final requests.
 regular accounts are according to',
   'AFRICA',
   JSON_PARSE('{"r_nations":[
      {"n_comment":" haggle. carefully final deposits detect slyly agai",
         "n_nationkey":0,
         "n_name":"ALGERIA"
```

```
      },
      {"n_comment":"ven packages wake quickly. regu",
         "n_nationkey":5,
         "n_name":"ETHIOPIA"
      },
      {"n_comment":" pending excuses haggle furiously deposits. pending, express pinto
 beans wake fluffily past t",
         "n_nationkey":14,
         "n_name":"KENYA"
      },
      {"n_comment":"rns. blithely bold courts among the closely regular packages use
 furiously bold platelets?",
         "n_nationkey":15,
         "n_name":"MOROCCO"
      },
      {"n_comment":"s. ironic, unusual asymptotes wake blithely r",
         "n_nationkey":16,
         "n_name":"MOZAMBIQUE"
      }
   ]
}'));
```

## Uso de COPY para cargar columnas SUPER en Amazon Redshift

En las secciones siguientes, puede obtener información acerca de diferentes formas de utilizar el comando COPY para cargar datos JSON en Amazon Redshift.

### Copia de datos de JSON y Avro

Mediante la compatibilidad con datos semiestructurados de Amazon Redshift, puede cargar un documento JSON sin fragmentar los atributos de sus estructuras JSON en varias columnas.

Amazon Redshift proporciona dos métodos para capturar documentos JSON mediante el uso de COPY, incluso con una estructura JSON que sea total o parcialmente desconocida:

1. Almacene los datos que derivan de un documento JSON en una sola columna de datos SUPER mediante la opción `noshred`. Este método es útil cuando el esquema no es conocido o se espera que cambie. Por lo tanto, este método hace que sea más fácil almacenar la tupla completa en una sola columna SUPER.
2. Fragmente el documento JSON en varias columnas de Amazon Redshift mediante la opción `auto` o `jsonpaths`. Los atributos pueden ser escalares de Amazon Redshift o valores SUPER.

Puede utilizar estas opciones con los formatos JSON o Avro.

El tamaño máximo de un objeto JSON antes de la fragmentación es de 4 MB.

## Copia de un documento JSON en una única columna de datos SUPER

Para copiar un documento JSON en una única columna de datos SUPER, cree una tabla con una sola columna de datos SUPER.

```
CREATE TABLE region_nations_noshred (rdata SUPER);
```

Copie los datos de Amazon S3 en la única columna de datos SUPER. Para capturar los datos de origen JSON en una sola columna de datos SUPER, especifique la opción `noshred` en la cláusula FORMAT JSON.

```
COPY region_nations_noshred FROM 's3://redshift-downloads/semistructured/tpch-nested/
data/json/region_nation'
REGION 'us-east-1' IAM_ROLE 'arn:aws:iam::xxxxxxxxxxxx:role/Redshift-S3'
FORMAT JSON 'noshred';
```

Después de que COPY haya capturado correctamente el JSON, su tabla tendrá una columna de datos SUPER `rdata` que contendrá los datos de todo el objeto JSON. Los datos capturados mantienen todas las propiedades de la jerarquía JSON. Sin embargo, las hojas se convierten en tipos escalares de Amazon Redshift para el procesamiento eficiente de la consulta.

Utilice la siguiente consulta para recuperar la cadena JSON original.

```
SELECT rdata FROM region_nations_noshred;
```

Cuando Amazon Redshift genera una columna de datos SUPER, se puede acceder a ella usando JDBC como cadena a través de la serialización JSON. Para obtener más información, consulte Serialización de datos JSON anidados y complejos.

## Copia de un documento JSON en varias columnas de datos SUPER

Puede fragmentar un documento JSON en varias columnas que pueden ser columnas de datos SUPER o tipos escalares de Amazon Redshift. Amazon Redshift propaga diferentes partes del objeto JSON a diferentes columnas.

```
CREATE TABLE region_nations
```

```
(
 r_regionkey smallint
 ,r_name varchar
 ,r_comment varchar
 ,r_nations super
 );
```

Para copiar los datos del ejemplo anterior en la tabla, especifique la opción AUTO en la cláusula FORMAT JSON para dividir el valor JSON en varias columnas. COPY hace coincidir los atributos JSON de nivel superior con los nombres de las columnas y permite que los valores anidados se capturen como valores SUPER, por ejemplo, matrices y objetos JSON.

```
COPY region_nations FROM 's3://redshift-downloads/semistructured/tpch-nested/data/json/
region_nation'
REGION 'us-east-1' IAM_ROLE 'arn:aws:iam::xxxxxxxxxxxx:role/Redshift-S3'
FORMAT JSON 'auto';
```

Cuando los nombres de atributos JSON están en mayúsculas y minúsculas combinadas, especifique la opción `auto ignorecase` en la cláusula FORMAT JSON. Para obtener más información acerca del comando COPY, consulte Carga desde datos JSON con la opción 'auto ignorecase'.

En algunos casos, hay una incompatibilidad entre los nombres de las columnas y los atributos JSON o el atributo que se va a cargar está anidado a más que un nivel de profundidad. Si es así, use un archivo `jsonpaths` para asignar manualmente los atributos JSON a las columnas de Amazon Redshift.

```
CREATE TABLE nations
(
 regionkey smallint
 ,name varchar
 ,comment super
 ,nations super
 );
```

Imagine que desea cargar datos en una tabla en la que los nombres de las columnas no coinciden con los atributos JSON. En el siguiente ejemplo, la tabla `nations` es de este tipo. Puede crear un archivo `jsonpaths` que asigne las rutas de los atributos a las columnas de la tabla por su posición en la matriz `jsonpaths`.

```
{"jsonpaths": [
```

```
       "$.r_regionkey",
       "$.r_name",
       "$.r_comment",
       "$.r_nations
    ]
}
```

La ubicación del archivo `jsonpaths` se utiliza como argumento para FORMAT JSON.

```
COPY nations FROM 's3://redshift-downloads/semistructured/tpch-nested/data/json/
region_nation'
REGION 'us-east-1' IAM_ROLE 'arn:aws:iam::xxxxxxxxxxxx:role/Redshift-S3'
FORMAT JSON 's3://redshift-downloads/semistructured/tpch-nested/data/jsonpaths/
nations_jsonpaths.json';
```

Utilice la siguiente consulta para acceder a la tabla que muestra los datos distribuidos en varias columnas. Las columnas de datos SUPER se imprimen mediante el formato JSON.

```
SELECT r_regionkey,r_name,r_comment,r_nations[0].n_nationkey FROM region_nations ORDER
 BY 1,2,3 LIMIT 1;
```

Los archivos jsonpaths asignan los campos del documento JSON a columnas de la tabla. Puede extraer columnas adicionales, como claves de distribución y ordenación, mientras sigue cargando el documento completo en una columna SUPER. La siguiente consulta carga el documento completo en la columna NATIONS. La columna `name` es la clave de ordenación y la columna `regionkey` es la clave de distribución.

```
CREATE TABLE nations_sorted (
    regionkey smallint,
    name varchar,
    nations super
) DISTKEY(regionkey) SORTKEY(name);
```

La raíz jsonpath “$” se asigna a la raíz del documento de la siguiente manera:

```
{"jsonpaths": [
       "$.r_regionkey",
       "$.r_name",
       "$"
    ]
```

```
}
```

La ubicación del archivo jsonpaths se utiliza como argumento para FORMAT JSON.

```
COPY nations_sorted FROM 's3://redshift-downloads/semistructured/tpch-nested/data/json/
region_nation'
REGION 'us-east-1' IAM_ROLE 'arn:aws:iam::xxxxxxxxxxxx:role/Redshift-S3'
FORMAT JSON 's3://redshift-downloads/semistructured/tpch-nested/data/jsonpaths/
nations_sorted_jsonpaths.json';
```

## Copia de datos desde texto y CSV

Amazon Redshift representa las columnas SUPER en formatos de texto y CSV como objetos JSON serializados. Se requiere un formato JSON válido para que las columnas SUPER se carguen con el tipo de información correcto. Elimine las comillas de los objetos, las matrices, los números, los valores booleanos y los valores nulos. Ajuste los valores de cadena con comillas dobles. Las columnas SUPER utilizan normas de escape estándar para los formatos de texto y CSV. Para CSV, a los delimitadores se les aplica un carácter de escape según el formato CSV estándar. Para el formato de texto, cuando el delimitador elegido también puede aparecer en un campo SUPER, utilice la opción ESCAPE durante las operaciones COPY y UNLOAD.

```
COPY region_nations FROM 's3://redshift-downloads/semistructured/tpch-nested/data/csv/
region_nation'
REGION 'us-east-1' IAM_ROLE 'arn:aws:iam::xxxxxxxxxxxx:role/Redshift-S3'
FORMAT CSV;
```

```
COPY region_nations FROM 's3://redshift-downloads/semistructured/tpch-nested/data/text/
region_nation'
REGION 'us-east-1' IAM_ROLE 'arn:aws:iam::xxxxxxxxxxxx:role/Redshift-S3'
DELIMITER ','
ESCAPE;
```

## Copia de datos de Parquet y ORC en formato de columnas

Si los datos semiestructurados o anidados ya están disponibles en formato Apache Parquet o Apache ORC, puede utilizar el comando COPY para capturar datos en Amazon Redshift.

La estructura de la tabla de Amazon Redshift debe coincidir con el número de columnas y con los tipos de datos de las columnas de los archivos Parquet u ORC. Si especifica SERIALIZETOJSON

en el comando COPY, puede cargar cualquier tipo de columna en el archivo que se alinee con una columna SUPER de la tabla como SUPER. Esto incluye tipos de estructura y matriz.

```
COPY region_nations FROM 's3://redshift-downloads/semistructured/tpch-nested/data/
parquet/region_nation'
REGION 'us-east-1' IAM_ROLE 'arn:aws:iam::xxxxxxxxxxxx:role/Redshift-S3'
FORMAT PARQUET SERIALIZETOJSON;
```

En el siguiente ejemplo, se utiliza un formato ORC.

```
COPY region_nations FROM 's3://redshift-downloads/semistructured/tpch-nested/data/orc/
region_nation'
IAM_ROLE 'arn:aws:iam::xxxxxxxxxxxx:role/Redshift-S3'
FORMAT ORC SERIALIZETOJSON;
```

Cuando los atributos de los tipos de datos de fecha u hora están en ORC, Amazon Redshift los convierte en varchar luego de codificarlos en SUPER.

## Descarga de datos semiestructurados

Con Amazon Redshift, puede exportar datos semiestructurados del clúster de Amazon Redshift a Amazon S3 en una variedad de formatos, incluidos texto, Apache Parquet, Apache ORC y Avro. En las siguientes secciones, se le guiará a través del proceso de configuración y ejecución de operaciones de descarga para los datos semiestructurados en Amazon Redshift.

CSV or text formats

Puede descargar tablas con columnas de datos SUPER en Amazon S3 en un formato de valores separados por comas (CSV) o de texto. Mediante una combinación de cláusulas de navegación y eliminación de anidamiento, Amazon Redshift descarga datos jerárquicos en formato de datos SUPER en Amazon S3 en formato CSV o de texto. Posteriormente, puede crear tablas externas para los datos descargados y consultarlas con Redshift Spectrum. Para obtener más información acerca del uso de UNLOAD y de los permisos de IAM necesarios, consulte UNLOAD.

Antes de ejecutar el siguiente ejemplo, rellene la tabla region_nations siguiendo los procesos de Carga de datos semiestructurados en Amazon Redshift. Para obtener información acerca de las tablas utilizadas en los ejemplos siguientes, consulte Conjunto de datos de muestra SUPER.

En el siguiente ejemplo, se descargan datos en Amazon S3.

```
UNLOAD ('SELECT * FROM region_nations')
TO 's3://xxxxxx/'
IAM_ROLE 'arn:aws:iam::xxxxxxxxxxxx:role/Redshift-S3-Write'
DELIMITER AS '|'
GZIP
ALLOWOVERWRITE;
```

A diferencia de otros tipos de datos en los que una cadena definida por el usuario representa un valor nulo, Amazon Redshift exporta las columnas de datos SUPER mediante el formato JSON y lo representa como nulo, de acuerdo con lo determinado por el formato JSON. Como resultado, las columnas de datos SUPER ignoran la opción NULL [AS] utilizada en los comandos UNLOAD.

Parquet format

Puede descargar tablas con columnas de datos SUPER en Amazon S3 en formato Parquet. Amazon Redshift representa las columnas SUPER en Parquet como el tipo de datos JSON. Esto permite representar datos semiestructurados en Parquet. Puede consultar estas columnas con Redshift Spectrum o volver a introducirlas en Amazon Redshift con el comando COPY. Para obtener más información acerca del uso de UNLOAD y de los permisos de IAM necesarios, consulte UNLOAD.

En el siguiente ejemplo, se descargan datos en Amazon S3 con formato Parquet.

```
UNLOAD ('SELECT * FROM region_nations')
TO 's3://xxxxxx/'
IAM_ROLE 'arn:aws:iam::xxxxxxxxxxxx:role/Redshift-S3-Write'
FORMAT PARQUET;
```

# Consulta de datos semiestructurados

Con Amazon Redshift, puede consultar y analizar datos semiestructurados, como JSON, Avro o Ion, junto con los datos estructurados. Los datos semiestructurados se refieren a los datos que tienen un esquema flexible, lo que permite estructuras jerárquicas o anidadas. En las siguientes secciones, se muestra la consulta de datos semiestructurados mediante el uso del soporte de Amazon Redshift para formatos de datos abiertos, lo que le permite obtener información valiosa de estructuras de datos complejas.

Amazon Redshift utiliza el lenguaje PartiQL para ofrecer acceso compatible con SQL a datos relacionales, semiestructurados y anidados.

PartiQL funciona con tipos dinámicos. Este enfoque permite filtrar, unir y agrupar de forma intuitiva la combinación de conjuntos de datos estructurados, semiestructurados y anidados. La sintaxis PartiQL utiliza notación con puntos y subíndice de matriz para la navegación de rutas cuando se accede a datos anidados. También habilita los elementos de la cláusula FROM para iterar sobre matrices y utilizarlos para las operaciones de eliminación de anidamiento. A continuación, se describen los diferentes patrones de consulta que combinan el uso del tipo de datos SUPER con la navegación, la anulación de dinamización, la eliminación de anidamiento o las uniones de las rutas y las matrices.

Para obtener información acerca de las tablas utilizadas en el ejemplo siguiente, incluido el esquema, consulte Conjunto de datos de muestra SUPER.

Temas


- Navegación
- Consultas de eliminación del anidamiento
- Anulación de la dinamización de los objetos
- Escritura dinámica
- Semántica laxa
- Tipos de introspección
- Ordenar por


## Navegación

Amazon Redshift utiliza PartiQL para permitir la navegación hacia matrices y estructuras mediante la notación […] con corchetes y puntos respectivamente. Además, puede mezclar la navegación en estructuras mediante la notación con puntos y matrices con la notación con corchetes. Por ejemplo, en el siguiente ejemplo, se supone que la columna de datos SUPER `c_orders` es una matriz con una estructura y que un atributo se denomina `o_orderkey`.

Para capturar datos en la tabla `customer_orders_lineitem`, ejecute el siguiente comando. Reemplace el rol de IAM con sus propias credenciales.

```
COPY customer_orders_lineitem FROM 's3://redshift-downloads/semistructured/tpch-nested/
data/json/customer_orders_lineitem'
REGION 'us-east-1' IAM_ROLE 'arn:aws:iam::xxxxxxxxxxxx:role/Redshift-S3'
FORMAT JSON 'auto';
```

```
SELECT c_orders[0].o_orderkey FROM customer_orders_lineitem;
```

Amazon Redshift también utiliza un alias de tablas como prefijo para la notación. En el siguiente ejemplo, se presenta la misma consulta que en el ejemplo anterior.

```
SELECT cust.c_orders[0].o_orderkey FROM customer_orders_lineitem AS cust;
```

Puede utilizar las notaciones con puntos y corchetes en todos los tipos de consultas, como las de filtrado, combinación y agregación. También puede utilizar estas notaciones en una consulta en la que por lo general hay referencias de columnas. En el siguiente ejemplo, se utiliza una instrucción SELECT que filtra los resultados.

```
SELECT count(*) FROM customer_orders_lineitem WHERE c_orders[0]. o_orderkey IS NOT
 NULL;
```

En el siguiente ejemplo, se utiliza la navegación con corchetes y puntos tanto en las cláusulas GROUP BY como ORDER BY.

```
SELECT c_orders[0].o_orderdate,
       c_orders[0].o_orderstatus,
       count(*)
FROM customer_orders_lineitem
WHERE c_orders[0].o_orderkey IS NOT NULL
GROUP BY c_orders[0].o_orderstatus,
         c_orders[0].o_orderdate
ORDER BY c_orders[0].o_orderdate;
```

## Consultas de eliminación del anidamiento

Para eliminar el anidamiento de las consultas, Amazon Redshift utiliza la sintaxis de PartiQL para iterar sobre las matrices SUPER. Para ello, navega por la matriz con la cláusula FROM de una consulta. Continuando con el ejemplo anterior, el siguiente ejemplo itera los valores de atributo de `c_orders`.

```
SELECT c.*, o FROM customer_orders_lineitem c, c.c_orders o;
```

La sintaxis de desanidamiento es una extensión de la cláusula FROM. En SQL estándar, la cláusula FROM `x (AS)` y significa que y itera cada tupla en relación con `x`. En este caso, `x` hace referencia

a una relación y y hace referencia a un alias de relación x. Del mismo modo, la sintaxis de PartiQL de eliminación de anidamiento con el elemento de cláusula FROM x (AS) y significa que y itera sobre cada valor (SUPER) en la expresión de matriz (SUPER) x. En este caso, x es una expresión SUPER y y es un alias de x.

El operando izquierdo también puede utilizar la notación con puntos y corchetes para la navegación normal. En el ejemplo anterior, customer_orders_lineitem c es la iteración sobre la tabla base customer_order_lineitem y c.c_orders o es la iteración sobre la matriz c.c_orders. Para iterar sobre el atributo o_lineitems, que es una matriz dentro de otra matriz, debe agregar varias cláusulas.

```
SELECT c.*, o, l FROM customer_orders_lineitem c, c.c_orders o, o.o_lineitems l;
```

Amazon Redshift también admite un índice de matrices cuando se itera sobre la matriz usando la palabra clave AT. La cláusula x AS y AT z itera sobre la matriz x y genera el campo z,, que es el índice de la matriz. En el siguiente ejemplo, se muestra cómo funciona un índice de matrices.

```
SELECT c_name,
       orders.o_orderkey AS orderkey,
       index AS orderkey_index
FROM customer_orders_lineitem c, c.c_orders AS orders AT index
ORDER BY orderkey_index;

c_name             | orderkey | orderkey_index
-------------------+----------+----------------
Customer#000008251 | 3020007  |        0
Customer#000009452 | 4043971  |        0
  (2 rows)
```

En el siguiente ejemplo se itera una matriz escalar.

```
CREATE TABLE bar AS SELECT json_parse('{"scalar_array": [1, 2.3, 45000000]}') AS data;

SELECT index, element FROM bar AS b, b.data.scalar_array AS element AT index;

 index | element
-------+----------
     0 | 1
     1 | 2.3
     2 | 45000000
```

```
(3 rows)
```

En el siguiente ejemplo se itera una matriz de varios niveles. En el ejemplo se utilizan varias cláusulas no anidadas para iterar en las matrices más internas. La matriz AS `f.multi_level_array` itera sobre `multi_level_array`. El elemento AS de la matriz representa la iteración sobre las matrices dentro de `multi_level_array`.

```
CREATE TABLE foo AS SELECT json_parse('[[1.1, 1.2], [2.1, 2.2], [3.1, 3.2]]') AS
 multi_level_array;

SELECT array, element FROM foo AS f, f.multi_level_array AS array, array AS element;

   array   | element
-----------+---------
 [1.1,1.2] | 1.1
 [1.1,1.2] | 1.2
 [2.1,2.2] | 2.1
 [2.1,2.2] | 2.2
 [3.1,3.2] | 3.1
 [3.1,3.2] | 3.2
(6 rows)
```

Para obtener más información acerca de la cláusula FROM, consulte Cláusula FROM.

## Anulación de la dinamización de los objetos

Para anular la dinamización del objeto, Amazon Redshift utiliza la sintaxis de PartiQL para iterar sobre objetos SUPER. Para ello, utiliza la cláusula FROM de una consulta con la palabra clave UNPIVOT. En este caso, la expresión es el objeto `c.c_orders[0]`. La consulta de ejemplo itera sobre cada atributo devuelto por el objeto.

```
SELECT attr as attribute_name, json_typeof(val) as value_type
FROM customer_orders_lineitem c, UNPIVOT c.c_orders[0] AS val AT attr
WHERE c_custkey = 9451;

 attribute_name  | value_type
-----------------+------------
 o_orderstatus   | string
 o_clerk         | string
 o_lineitems     | array
 o_orderdate     | string
```

```
 o_shippriority  | number
 o_totalprice    | number
 o_orderkey      | number
 o_comment       | string
 o_orderpriority | string
(9 rows)
```

Al igual que para eliminar el anidamiento, la sintaxis para anular la dinamización también es una extensión de la cláusula FROM. La diferencia es que la sintaxis para anular la dinamización utiliza la palabra clave UNPIVOT para indicar que está iterando sobre un objeto en lugar de una matriz. Utiliza AS `value_alias` para iterar sobre todos los valores dentro de un objeto y utiliza AT `attribute_alias` para iterar sobre todos los atributos. Considere el siguiente fragmento de sintaxis:

```
UNPIVOT expression AS value_alias [ AT attribute_alias ]
```

Amazon Redshift admite la anulación de la dinamización de objetos y la eliminación del anidamiento de matrices en una sola cláusula FROM de la siguiente manera:

```
SELECT attr as attribute_name, val as object_value
FROM customer_orders_lineitem c, c.c_orders AS o, UNPIVOT o AS val AT attr
WHERE c_custkey = 9451;
```

Cuando se utiliza la anulación de la dinamización de objetos, Amazon Redshift no admite la anulación correlacionada. En concreto, supongamos que tiene un caso en el que hay varios ejemplos de anulación de la dinamización en diferentes niveles de consulta y la anulación de la dinamización interna hace referencia al exterior. Amazon Redshift no admite este tipo de anulación de la dinamización múltiple.

Para obtener más información acerca de la cláusula FROM, consulte Cláusula FROM. Para ver ejemplos que muestran cómo realizar consultas de datos estructurados con PIVOT y UNPIVOT, consulte Ejemplos de PIVOT y UNPIVOT.

## Escritura dinámica

La escritura dinámica no requiere la conversión explícita de los datos que se extraen de las rutas con puntos y corchetes. Amazon Redshift utiliza la escritura dinámica para procesar datos SUPER sin esquema sin necesidad de declarar los tipos de datos antes de utilizarlos en la consulta. La escritura

dinámica utiliza los resultados de navegar por columnas de datos SUPER sin tener que convertirlos explícitamente en tipos de Amazon Redshift. La escritura dinámica es más útil en las uniones y las cláusulas GROUP BY. En el siguiente ejemplo, se utiliza una instrucción SELECT que no requiere ninguna conversión explícita de las expresiones con puntos y corchetes en los tipos habituales de Amazon Redshift. Para obtener más información acerca de la compatibilidad y la conversión de tipos, consulte Conversión y compatibilidad de tipos.

```
SELECT c_orders[0].o_orderkey
FROM customer_orders_lineitem
WHERE c_orders[0].o_orderstatus = 'P';
```

El signo de igualdad en esta consulta se evalúa como `true` cuando c_orders[0].o_orderstatus es la cadena “P”. En todos los demás casos, el signo de igualdad se evalúa como `false`, incluidos los casos en los que los argumentos de la igualdad son tipos diferentes.

## Escritura dinámica y estática

Sin usar la escritura dinámica, no se puede determinar si c_orders[0].o_orderstatus es una cadena, un entero o una estructura. Solo puede determinar que c_orders[0].o_orderstatus es un tipo de datos SUPER, que puede ser un escalar, una matriz o una estructura de Amazon Redshift. El tipo estático de c_orders[0].o_orderstatus es un tipo de datos SUPER. De forma convencional, un tipo es implícitamente un tipo estático en SQL.

Amazon Redshift utiliza la escritura dinámica para el procesamiento de datos sin esquema. Cuando la consulta evalúa los datos, c_orders[0].o_orderstatus resulta ser un tipo específico. Por ejemplo, evaluar c_orders[0].o_orderstatus en el primer registro de customer_orders_lineitem puede resultar en un entero. La evaluación en el segundo registro puede resultar en una cadena. Estos son los tipos dinámicos de la expresión.

Cuando se utiliza un operador o una función SQL con expresiones con puntos y corchetes que tienen tipos dinámicos, Amazon Redshift produce resultados similares a utilizar el operador o la función SQL estándar con los respectivos tipos estáticos. En este ejemplo, cuando el tipo dinámico de la expresión de ruta es una cadena, la comparación con la cadena “P” es significativa. Siempre que el tipo dinámico de c_orders[0].o_orderstatus sea cualquier otro tipo de datos, excepto que sea una cadena, la igualdad devolverá el resultado false. Otras funciones devuelven valores nulos cuando se usan argumentos mal escritos.

En el siguiente ejemplo, se la consulta anterior se escribe con escritura estática:

```
SELECT c_custkey
FROM customer_orders_lineitem
WHERE CASE WHEN JSON_TYPEOF(c_orders[0].o_orderstatus) = 'string'
           THEN c_orders[0].o_orderstatus::VARCHAR = 'P'
           ELSE FALSE END;
```

Tenga en cuenta la siguiente distinción entre predicados de igualdad y predicados de comparación. En el ejemplo anterior, si el predicado de igualdad se reemplaza con un predicado de menor que o igual, la semántica produce un valor nulo en vez de uno false.

```
SELECT c_orders[0]. o_orderkey
FROM customer_orders_lineitem
WHERE c_orders[0].o_orderstatus <= 'P';
```

En este ejemplo, si c_orders[0].o_orderstatus es una cadena, Amazon Redshift devuelve el valor true si es alfabéticamente igual o menor que “P”. Amazon Redshift devuelve un valor false si es alfabéticamente mayor que “P”. Sin embargo, si c_orders[0].o_orderstatus no es una cadena, Amazon Redshift devuelve un valor “null” (nulo) debido a que Amazon Redshift no puede comparar valores de diferentes tipos, como se muestra en la consulta siguiente:

```
SELECT c_custkey
FROM customer_orders_lineitem
WHERE CASE WHEN JSON_TYPEOF(c_orders[0].o_orderstatus) = 'string'
           THEN c_orders[0].o_orderstatus::VARCHAR <= 'P'
           ELSE NULL END;
```

La escritura dinámica no efectúa exclusiones de las comparaciones de tipos que son mínimamente comparables. Por ejemplo, puede convertir los tipos escalares CHAR y VARCHAR de Amazon Redshift en SUPER. Estos tipos son comparables como cadenas, lo que incluye ignorar los caracteres de espacio en blanco finales similares a los tipos CHAR y VARCHAR de Amazon Redshift. De manera similar, los valores enteros, decimales y de coma flotante son comparables como valores SUPER. Específicamente para las columnas decimales, cada valor también puede tener una escala diferente. Amazon Redshift sigue considerándolos como tipos dinámicos.

Amazon Redshift también admite la igualdad en los objetos y las matrices que se evalúan como de igualdad profunda, tales como la evaluación profunda de objetos o matrices y la comparación de todos los atributos. Utilice la igualdad profunda con precaución, ya que el proceso de realizar la igualdad profunda puede llevar mucho tiempo.

## Uso de la escritura dinámica para las uniones

Para las uniones, la escritura dinámica combina automáticamente los valores con diferentes tipos dinámicos, sin tener que realizar un largo análisis de CASE WHEN para averiguar qué tipos de datos pueden aparecer. Por ejemplo, suponga que, con el tiempo, su organización cambió el formato que utilizaba para las claves de partes.

Las claves iniciales de partes de enteros emitidas se reemplazan por claves de partes de cadenas, como “A55”, y, luego, se reemplazan de nuevo por claves de partes de matrices, como [‘X’, 10], que combina una cadena y un número. Amazon Redshift no tiene que realizar un análisis prolongado de casos sobre las claves de partes y puede utilizar uniones, como se muestra en el siguiente ejemplo.

```
SELECT c.c_name
    ,l.l_extendedprice
    ,l.l_discount
FROM customer_orders_lineitem c
    ,c.c_orders o
    ,o.o_lineitems l
    ,supplier_partsupp s
    ,s.s_partsupps ps
WHERE l.l_partkey = ps.ps_partkey
AND c.c_nationkey = s.s_nationkey
ORDER BY c.c_name;
```

En el siguiente ejemplo, se muestra cuán compleja e ineficiente puede ser la misma consulta si no se usa la escritura dinámica:

```
SELECT c.c_name
    ,l.l_extendedprice
    ,l.l_discount
FROM customer_orders_lineitem c
    ,c.c_orders o
    ,o.o_lineitems l
    ,supplier_partsupp s
    ,s.s_partsupps ps
WHERE CASE WHEN IS_INTEGER(l.l_partkey) AND IS_INTEGER(ps.ps_partkey)
           THEN l.l_partkey::integer = ps.ps_partkey::integer
           WHEN IS_VARCHAR(l.l_partkey) AND IS_VARCHAR(ps.ps_partkey)
           THEN l.l_partkey::varchar = ps.ps_partkey::varchar
           WHEN IS_ARRAY(l.l_partkey) AND IS_ARRAY(ps.ps_partkey)
                AND IS_VARCHAR(l.l_partkey[0]) AND IS_VARCHAR(ps.ps_partkey[0])
```

```
                AND IS_INTEGER(l.l_partkey[1]) AND IS_INTEGER(ps.ps_partkey[1])
           THEN l.l_partkey[0]::varchar = ps.ps_partkey[0]::varchar
                AND l.l_partkey[1]::integer = ps.ps_partkey[1]::integer
           ELSE FALSE END
AND c.c_nationkey = s.s_nationkey
ORDER BY c.c_name;
```

## Semántica laxa

De manera predeterminada, las operaciones de navegación en los valores SUPER devuelven valores nulos en lugar de devolver un error cuando la navegación no es válida. La navegación por objetos no es válida si el valor SUPER no es un objeto o si el valor SUPER es un objeto, pero no contiene el nombre del atributo utilizado en la consulta. Por ejemplo, la siguiente consulta tiene acceso a un nombre de atributo no válido de la columna de datos SUPER cdata:

```
SELECT c.c_orders.something FROM customer_orders_lineitem c;
```

La navegación por matrices devuelve el valor nulo si el valor SUPER no es una matriz o si el índice de matriz está fuera de los límites. La siguiente consulta devuelve el valor nulo porque c_orders[1][1] está fuera de los límites.

```
SELECT c.c_orders[1][1] FROM customer_orders_lineitem c;
```

La semántica laxa es especialmente útil cuando se utiliza la escritura dinámica para convertir un valor SUPER. Convertir un valor SUPER en un tipo incorrecto devuelve el valor nulo en lugar de un error si la conversión no es válida. Por ejemplo, la siguiente consulta devuelve el valor nulo porque no puede convertir el valor de cadena “Good” del atributo de objeto o_orderstatus en INTEGER. Amazon Redshift devuelve un error para una conversión de VARCHAR en INTEGER, pero no para una conversión SUPER.

```
SELECT c.c_orders.o_orderstatus::integer FROM customer_orders_lineitem c;
```

## Tipos de introspección

Las columnas de datos SUPER admiten funciones de inspección que devuelven el tipo dinámico y otra información del tipo relativa al valor SUPER. El ejemplo más común es la función escalar JSON_TYPEOF que devuelve un VARCHAR con valores booleanos, de números, de cadenas, de

objetos, de matrices o nulos, dependiendo del tipo dinámico del valor SUPER. Amazon Redshift admite las siguientes funciones booleanas para las columnas de datos SUPER:

- DECIMAL_PRECISION
- DECIMAL_SCALE
- IS_ARRAY
- IS_BIGINT
- IS_CHAR
- IS_DECIMAL
- IS_FLOAT
- IS_INTEGER
- IS_OBJECT
- IS_SCALAR
- IS_SMALLINT
- IS_VARCHAR
- JSON_TYPEOF

Todas estas funciones devuelven un valor false si el valor de entrada es nulo. IS_SCALAR, IS_OBJECT e IS_ARRAY son mutuamente excluyentes y cubren todos los valores posibles, excepto los nulos.

Para inferir los tipos correspondientes a los datos, Amazon Redshift utiliza la función JSON_TYPEOF, que devuelve el tipo (el nivel superior) del valor SUPER, como se muestra en el siguiente ejemplo:

```
SELECT JSON_TYPEOF(r_nations) FROM region_nations;
 json_typeof
 -------------
 array
(1 row)
```

```
SELECT JSON_TYPEOF(r_nations[0].n_nationkey) FROM region_nations;
 json_typeof
 -------------
```

```
 number
```

Amazon Redshift ve esto como una única cadena larga, similar a insertar este valor en una columna VARCHAR en lugar de una SUPER. Como la columna es SUPER, la cadena única sigue siendo un valor SUPER válido y la diferencia se observa en JSON_TYPEOF:

```
SELECT IS_VARCHAR(r_nations[0].n_name) FROM region_nations;
 is_varchar
-------------
 true
(1 row)
```

```
SELECT r_nations[4].n_name FROM region_nations
WHERE CASE WHEN IS_INTEGER(r_nations[4].n_nationkey)
           THEN r_nations[4].n_nationkey::INTEGER = 15
           ELSE false END;
```

## Ordenar por

Amazon Redshift no define comparaciones SUPER entre valores con diferentes tipos dinámicos. El valor SUPER que es una cadena no es más pequeño ni más grande que un valor SUPER que es un número. Para utilizar cláusulas ORDER BY con columnas SUPER, Amazon Redshift define un orden total entre los diferentes tipos que se deben observar cuando Amazon Redshift clasifica los valores SUPER mediante cláusulas ORDER BY. El orden entre los tipos dinámicos es booleano, de número, de cadena, de matriz, de objeto. En el siguiente ejemplo, se muestran los órdenes de los diferentes tipos:

```
INSERT INTO region_nations VALUES
(100,'name1','comment1','AWS'),
(200,'name2','comment2',1),
(300,'name3','comment3',ARRAY(1, 'abc', null)),
(400,'name4','comment4',-2.5),
(500,'name5','comment5','Amazon');

SELECT r_nations FROM region_nations order by r_nations;

r_nations
----------------
 -2.5
 1
```

```
 "Amazon"
 "AWS"
 [1,"abc",null]
(5 rows)
```

Para obtener más información acerca de la cláusula ORDER BY, consulte Cláusula ORDER BY.

# Operadores y funciones

Con Amazon Redshift, puede realizar análisis avanzados en conjuntos de datos de gran tamaño utilizando datos SUPER mediante operadores y funciones. Los operadores y las funciones de los datos SUPER son constructos de SQL que permiten el análisis y la manipulación complejos de los datos semiestructurados almacenados en las tablas de Amazon Redshift.

En las siguientes secciones se describen la sintaxis, los ejemplos y las prácticas recomendadas para utilizar operadores y funciones para los datos SUPER en Amazon Redshift a fin de aprovechar todo el potencial de los datos semiestructurados.

## Operadores aritméticos

Los valores SUPER admiten todos los operadores aritméticos básicos +, -, *, /, % utilizando la escritura dinámica. El tipo resultante de la operación sigue siendo SUPER. Para todos los operadores, excepto para el operador binario +, los operandos de entrada deben ser números. De lo contrario, Amazon Redshift devolverá un valor nulo. La distinción entre valores decimales y de coma flotante se conserva cuando Amazon Redshift ejecuta estos operadores y el tipo dinámico no cambia. Sin embargo, la escala decimal cambia cuando se utilizan multiplicaciones y divisiones. Los desbordamientos aritméticos siguen causando errores de consulta; no se cambian a valores nulos. El operador binario + lleva a cabo la adición si las entradas son números o la concatenación si las entradas son cadenas. Si un operando es una cadena y el otro operando es un número, el resultado será nulo. Los operadores de prefijo unario + y - devuelven valores nulos si el valor SUPER no es un número, como se muestra en el siguiente ejemplo:

```
SELECT (c_orders[0]. o_orderkey + 0.5) * c_orders[0]. o_orderkey / 10 AS math FROM
 customer_orders_lineitem;
            math
----------------------------
 1757958232200.1500
(1 row)
```

La escritura dinámica permite que los valores decimales en SUPER tengan diferentes escalas. Amazon Redshift trata los valores decimales como si fueran diferentes tipos estáticos y permite todas las operaciones matemáticas. Amazon Redshift calcula la escala resultante de forma dinámica en función de las escalas de los operandos. Si uno de los operandos es un número de coma flotante, Amazon Redshift promueve el otro operando a un número de coma flotante y genera el resultado como un número de coma flotante.

## Funciones aritméticas

Amazon Redshift admite las siguientes funciones aritméticas para las columnas SUPER. Ellas devolverán un valor nulo si la entrada no es un número:

- FLOOR Para obtener más información, consulte Función FLOOR.
- CEIL y CEILING Para obtener más información, consulte Función CEILING (o CEIL).
- ROUND Para obtener más información, consulte Función ROUND.
- TRUNC Para obtener más información, consulte Función TRUNC.
- ABS Para obtener más información, consulte Función ABS.

En el siguiente ejemplo, se utilizan funciones aritméticas para consultar datos:

```
SELECT x, FLOOR(x), CEIL(x), ROUND(x)
FROM (
    SELECT (c_orders[0]. o_orderkey + 0.5) * c_orders[0].o_orderkey / 10 AS x
    FROM customer_orders_lineitem
    );

         x          |     floor     |     ceil      |     round
--------------------+---------------+---------------+---------------
 1389636795898.0500  | 1389636795898  | 1389636795899  | 1389636795898
```

La función ABS conserva la escala del decimal de entrada durante las funciones FLOOR, CEIL. La función ROUND elimina la escala del decimal de entrada.

## Funciones de matriz

Amazon Redshift admite las siguientes funciones de composición y utilidad de matriz: array_concat, subarray, array_flatten, get_array_length y split_to_array.

Puede construir matrices SUPER a partir de valores de los tipos de datos de Amazon Redshift mediante la función ARRAY, incluidos otros valores SUPER. En el siguiente ejemplo, se utiliza la función variádica ARRAY:

```
SELECT ARRAY(1, c.c_custkey, NULL, c.c_name, 'abc') FROM customer_orders_lineitem c;
                               array
 ------------------------------------------------------
[1,8401,null,""Customer#000008401"",""abc""]
[1,9452,null,""Customer#000009452"",""abc""]
[1,9451,null,""Customer#000009451"",""abc""]
[1,8251,null,""Customer#000008251"",""abc""]
[1,5851,null,""Customer#000005851"",""abc""]
(5 rows)
```

En el siguiente ejemplo, se utiliza la concatenación de matrices con la función ARRAY_CONCAT:

```
SELECT ARRAY_CONCAT(JSON_PARSE('[10001,10002]'),JSON_PARSE('[10003,10004]'));

             array_concat
-----------------------------------
 [10001,10002,10003,10004]
(1 row)
```

En el siguiente ejemplo, se utiliza la manipulación de matrices con la función SUBARRAY, que devuelve un subconjunto de la matriz de entrada.

```
SELECT SUBARRAY(ARRAY('a', 'b', 'c', 'd', 'e', 'f'), 2, 3);

   subarray
---------------
 ["c","d","e"]
(1 row))
```

En el siguiente ejemplo, se fusionan varios niveles de matrices en una sola matriz con ARRAY_FLATTEN:

```
SELECT x, ARRAY_FLATTEN(x) FROM (SELECT ARRAY(1, ARRAY(2, ARRAY(3, ARRAY())))) AS x);

     x           | array_flatten
 ----------------+---------------
```

```
 [1,[2,[3,[]]]]  | [1,2,3]
(1 row)
```

Las funciones de matriz ARRAY_CONCAT y ARRAY_FLATTEN utilizan reglas de escritura dinámica. Si la entrada no es una matriz, devuelven un valor nulo en lugar de un error. La función GET_ARRAY_LENGTH devuelve la longitud de una matriz SUPER dada una ruta de objeto o matriz.

```
SELECT c_name
FROM customer_orders_lineitem
WHERE GET_ARRAY_LENGTH(c_orders) = (
    SELECT MAX(GET_ARRAY_LENGTH(c_orders))
    FROM customer_orders_lineitem
    );
```

En el siguiente ejemplo, se divide una cadena en una matriz de cadenas usando SPLIT_TO_ARRAY. La función utiliza un delimitador como parámetro opcional. Si ningún delimitador está ausente, entonces el valor predeterminado es la coma.

```
SELECT SPLIT_TO_ARRAY('12|345|6789', '|');

   split_to_array
---------------------
 ["12","345","6789"]
(1 row)
```

# Configuraciones SUPER

Puede configurar los datos SUPER para situaciones específicas. En las siguientes secciones se proporcionan detalles sobre cómo elegir y aplicar las configuraciones SUPER adecuadas en función de los requisitos de formato de datos.

Temas

- Modos laxo y estricto de SUPER
- Acceso a campos JSON con nombres de campo o atributos en mayúsculas y con mayúsculas y minúsculas combinadas
- Opciones de análisis de SUPER

## Modos laxo y estricto de SUPER

Cuando consulta datos SUPER, la expresión de ruta puede no concordar con la estructura real de los datos SUPER. Si intenta acceder a un miembro inexistente de un objeto o un elemento de una matriz, Amazon Redshift devuelve un valor NULL cuando la consulta se ejecuta en el modo laxo predeterminado. Si ejecuta la consulta en el modo estricto, Amazon Redshift devuelve un error. Los siguientes parámetros de sesión se pueden configurar para activar o desactivar el modo laxo.

En el siguiente ejemplo, se utilizan parámetros de sesión para habilitar el modo laxo.

```
SET navigate_super_null_on_error=ON;  --default lax mode for navigation

SET cast_super_null_on_error=ON;  --default lax mode for casting

SET parse_super_null_on_error=OFF;  --default strict mode for ingestion
```

## Acceso a campos JSON con nombres de campo o atributos en mayúsculas y con mayúsculas y minúsculas combinadas

Cuando los nombres de los atributos JSON están en mayúsculas o en mayúsculas y minúsculas combinadas, debe poder navegar por las estructuras de tipo SUPER con distinción entre mayúsculas y minúsculas. Para ello, puede configurar `enable_case_sensitive_identifier` en TRUE e incluir entre comillas dobles los nombres de atributo en mayúsculas y en mayúsculas y minúsculas combinadas. También puede configurar `enable_case_sensitive_super_attribute` en TRUE. En este caso, puede utilizar nombres de atributo en mayúsculas y en mayúsculas y minúsculas combinadas en las consultas sin incluirlos entre comillas dobles.

En el siguiente ejemplo, se ilustra cómo establecer `enable_case_sensitive_identifier` para consultar datos.

```
SET enable_case_sensitive_identifier to TRUE;

-- Accessing JSON attribute names with uppercase and mixedcase names
SELECT json_table.data."ITEMS"."Name",
       json_table.data."price"
FROM
  (SELECT json_parse('{"ITEMS":{"Name":"TV"}, "price": 345}') AS data) AS json_table;

 Name | price
------+-------
```

```
 "TV" | 345
(1 row)

RESET enable_case_sensitive_identifier;

-- After resetting the above configuration, the following query accessing JSON
 attribute names with uppercase and mixedcase names should return null (if in lax
 mode).
SELECT json_table.data."ITEMS"."Name",
       json_table.data."price"
FROM
  (SELECT json_parse('{"ITEMS":{"Name":"TV"}, "price": 345}') AS data) AS json_table;

 name | price
------+-------
      | 345
(1 row)
```

En el siguiente ejemplo, se ilustra cómo establecer `enable_case_sensitive_super_attribute` para consultar datos.

```
SET enable_case_sensitive_super_attribute to TRUE;
 -- Accessing JSON attribute names with uppercase and mixedcase names

SELECT json_table.data.ITEMS.Name,
       json_table.data.price
FROM
  (SELECT json_parse('{"ITEMS":{"Name":"TV"}, "price": 345}') AS data) AS json_table;

 name | price
 -----+-------
 "TV" | 345
(1 row)

 RESET enable_case_sensitive_super_attribute;

 -- After resetting enable_case_sensitive_super_attribute, the query now returns NULL
 for ITEMS.Name (if in lax mode).

SELECT json_table.data.ITEMS.Name,
       json_table.data.price
FROM
  (SELECT json_parse('{"ITEMS":{"Name":"TV"}, "price": 345}') AS data) AS json_table;
```

```
 name | price
 -----+-------
      | 345
(1 row)
```

## Opciones de análisis de SUPER

Cuando utiliza la función JSON_PARSE para analizar cadenas JSON en valores SUPER, se aplican determinadas restricciones:

- El mismo nombre de atributo no puede aparecer en el mismo objeto, pero puede aparecer en un objeto anidado. La opción de configuración `json_parse_dedup_attributes` permite a JSON_PARSE conservar solo la última aparición de atributos duplicados en lugar de devolver un error.
- Los valores de cadena no pueden superar el tamaño varchar máximo del sistema de 65535 bytes. La opción de configuración `json_parse_truncate_strings` permite a JSON_PARSE() truncar automáticamente cadenas que superan este límite sin devolver un error. Este comportamiento solo afecta a los valores de cadena y no a los nombres de atributos.

Para obtener más información acerca de la función JSON_PARSE, consulte Función JSON_PARSE.

En el siguiente ejemplo se muestra cómo configurar la opción `json_parse_dedup_attributes` al comportamiento predeterminado de devolver un error para atributos duplicados.

```
SET json_parse_dedup_attributes=OFF;  --default behavior of returning error instead of
 de-duplicating attributes
```

En el siguiente ejemplo se muestra cómo configurar la opción `json_parse_truncate_strings` para el comportamiento predeterminado de devolver un error para cadenas que ya no tienen este límite.

```
SET json_parse_truncate_strings=OFF;  --default behavior of returning error instead of
 truncating strings
```

# Limitaciones

Con Amazon Redshift, puede trabajar con el tipo de datos SUPER para almacenar y consultar datos semiestructurados como JSON, Avro o Ion. Las limitaciones de los tipos de datos SUPER se refieren a las restricciones y los límites cuando se utiliza este tipo de datos en Amazon Redshift. En las siguientes secciones se proporcionan detalles sobre las limitaciones específicas del tipo de datos SUPER, como el tamaño máximo, los niveles de anidación y los tipos de datos admitidos en los datos semiestructurados.

- No se pueden definir columnas SUPER como clave de distribución ni de ordenación.
- Un objeto SUPER individual puede almacenar hasta 16 MB de datos.
- La profundidad máxima de anidación para matrices y estructuras del tipo de datos SUPER es 1000.
- Un valor individual dentro de un objeto SUPER se limita a la longitud máxima del tipo de Amazon Redshift correspondiente. Por ejemplo, un único valor de cadena cargado en SUPER se limita a la longitud máxima de VARCHAR de 65 535 bytes.
- No se pueden realizar operaciones parciales de actualización ni transformación en las columnas SUPER.
- No se puede usar el tipo de datos SUPER ni su alias en uniones correctas ni uniones externas completas.
- El tipo de datos SUPER no admite XML como formato de serialización entrante ni saliente.
- En la cláusula FROM de una subconsulta (correlacionada o no) que referencia una variable de tabla para eliminar el anidamiento, la consulta solo puede hacer referencia a su tabla principal y no a otras tablas.
- No se puede eliminar el anidamiento ni anular la dinamización de objetos en columnas SUPER generadas por una expresión de tabla común recursiva (`WITH RECURSIVE`). Para obtener más información sobre las CTE recursivas, consulte. Expresiones de tabla comunes recursivas
- Limitaciones de conversión

  Los valores SUPER se pueden convertir en otros tipos de datos y a partir de otros tipos de datos, con las siguientes excepciones:

  - Amazon Redshift no diferencia entre números enteros y decimales de la escala 0.
  - Si la escala no es cero, el tipo de datos SUPER tiene el mismo comportamiento que otros tipos de datos de Amazon Redshift, excepto que Amazon Redshift convierte los errores relacionados con el tipo SUPER en valores nulos, como se muestra en el siguiente ejemplo.

```
SELECT 5::bool;
 bool
-------
 True
(1 row)

SELECT 5::decimal::bool;
ERROR:  cannot cast type numeric to boolean

SELECT 5::super::bool;
 bool
-------
 True
(1 row)

SELECT 5.0::bool;
ERROR:  cannot cast type numeric to boolean

SELECT 5.0::super::bool;
 bool
-------
(1 row)
```

- Amazon Redshift no convierte los tipos de fecha y hora en tipo de datos SUPER. Amazon Redshift solo puede convertir el tipo de datos SUPER en tipos de datos de fecha y hora, como se muestra en el siguiente ejemplo.

```
SELECT o.o_orderdate FROM customer_orders_lineitem c,c.c_orders o;
  order_date
------------------
 "2001-09-08"
(1 row)


SELECT JSON_TYPEOF(o.o_orderdate) FROM customer_orders_lineitem c,c.c_orders o;
 json_typeof
-----------------
 string
(1 row)


SELECT o.o_orderdate::date FROM customer_orders_lineitem c,c.c_orders o;
```

```
 order_date
----------------
 2001-09-08
(1 row)


--date/time cannot be cast to super
SELECT '2019-09-09'::date::super;
ERROR:  cannot cast type date to super
```

- La conversión de valores no escalares (objeto y matriz) en cadenas devuelve el valor NULL. Para serializar correctamente estos valores no escalares, no los convierta. En su lugar, utilice `json_serialize` para convertir los valores no escalares. La función `json_serialize` devuelve un varchar. Por lo general, no es necesario convertir valores no escalares en varchar, ya que Amazon Redshift lleva a cabo serializaciones de forma implícita, como se muestra en el primer ejemplo siguiente.

```
SELECT r_nations FROM region_nations WHERE r_regionkey=300;
   r_nations
----------------
 [1,"abc",null]
(1 row)

SELECT r_nations::varchar FROM region_nations WHERE r_regionkey=300;
 r_nations
-------------
(1 row)

SELECT JSON_SERIALIZE(r_nations) FROM region_nations WHERE r_regionkey=300;
 json_serialize
-----------------
 [1,"abc",null]
(1 row)
```

- En el caso de las bases de datos que no distinguen mayúsculas de minúsculas, Amazon Redshift no admite el tipo de datos SUPER. En el caso de las columnas que no distinguen mayúsculas de minúsculas, Amazon Redshift no las convierte en el tipo SUPER. Por lo tanto, Amazon Redshift no admite que las columnas SUPER interactúen con columnas que no distinguen mayúsculas de minúsculas y que desencadenan la conversión.

- Amazon Redshift no admite las funciones volátiles, como RANDOM ( ) o TIMEOFDAY ( ), en las subconsultas que eliminan el anidamiento de una tabla externa o un lado izquierdo (LHS) de funciones IN con dichas subconsultas.

# Tipo de datos SUPER y vistas materializadas

Con Amazon Redshift, puede utilizar el tipo de datos SUPER para mejorar el rendimiento y la flexibilidad de las vistas materializadas. El tipo de datos SUPER le permite almacenar un superconjunto de columnas de las tablas base en una vista materializada, lo que le permite consultar la vista materializada directamente sin unir las tablas base. En las siguientes secciones, se muestra cómo crear y utilizar vistas materializadas con el tipo de datos SUPER en Amazon Redshift.

Amazon Redshift amplía la capacidad de las vistas materializadas para trabajar con el tipo de datos SUPER y con PartiQL. Las consultas SQL y PartiQL se pueden calcular previamente mediante vistas materializadas progresivas. Para obtener más información acerca de las vistas materializadas, consulte Vistas materializadas en Amazon Redshift.

Una vez que haya almacenado los datos sin esquemas y semiestructurados como SUPER, puede utilizar vistas materializadas de PartiQL para realizar una introspección de los datos y fragmentarlos en vistas materializadas.

## Aceleración de consultas PartiQL

Puede utilizar vistas materializadas para acelerar las consultas PartiQL que navegan por los datos jerárquicos de las columnas SUPER o eliminan su anidamiento. Cree una o más vistas materializadas para fragmentar los valores SUPER en varias columnas y utilice la organización en columnas de las consultas analíticas de Amazon Redshift. En consecuencia, las consultas utilizan las vistas materializadas.

La vista materializada esencialmente extrae y normaliza los datos anidados. El nivel de normalización depende de cuánto esfuerzo destine a convertir los datos SUPER en datos convencionales divididos en columnas.

### Fragmentación en columnas SUPER con vistas materializadas

Con Amazon Redshift, puede mejorar el rendimiento de las consultas y reducir los requisitos de almacenamiento al triturar los datos en columnas SUPER mediante vistas materializadas. La trituración se refiere al proceso de dividir tipos de datos complejos, como JSON o XML semiestructurados, en columnas más pequeñas y planas. Las columnas SUPER son una forma

especializada de almacenamiento en columnas optimizada para escanear rápidamente los datos triturados.

En las siguientes secciones se describen los pasos y las consideraciones para triturar datos en columnas SUPER mediante vistas materializadas en Amazon Redshift.

En el siguiente ejemplo, se muestra una vista materializada que fragmenta los datos anidados, mientras las columnas resultantes siguen siendo del tipo de datos SUPER.

```
SELECT c.c_name, o.o_orderstatus
FROM customer_orders_lineitem c, c.c_orders o;
```

En el siguiente ejemplo, se muestra una vista materializada que crea columnas escalares convencionales de Amazon Redshift a partir de los datos fragmentados.

```
SELECT c.c_name, c.c_orders[0].o_totalprice
FROM customer_orders_lineitem c;
```

Puede crear una única vista materializada super_mv para acelerar ambas consultas.

Para responder la primera consulta, debe materializar el atributo o_orderstatus. Puede omitir el atributo c_name porque no implica la navegación anidada ni la eliminación del anidamiento. También debe incluir en la vista materializada el atributo c_custkey de customer_orders_lineitem para poder unir la tabla base con la vista materializada.

Para responder la segunda consulta, también debe materializar el atributo o_totalprice y el índice de matriz o_idx de c_orders. Por consiguiente, puede acceder al índice 0 de c_orders.

```
CREATE MATERIALIZED VIEW super_mv distkey(c_custkey) sortkey(c_custkey) AS (
  SELECT c_custkey, o.o_orderstatus, o.o_totalprice, o_idx
  FROM customer_orders_lineitem c, c.c_orders o AT o_idx
);
```

Los atributos o_orderstatus y o_totalprice de la vista materializada super_mv son de tipo SUPER.

La vista materializada super_mv se actualizará progresivamente luego de los cambios que se realicen en la tabla base customer_orders_lineitem.

```
REFRESH MATERIALIZED VIEW super_mv;
INFO: Materialized view super_mv was incrementally updated successfully.
```

Para volver a escribir la primera consulta PartiQL como una consulta SQL normal, una customer_orders_lineitem con super_mv de la siguiente manera.

```
SELECT c.c_name, v.o_orderstatus
FROM customer_orders_lineitem c
JOIN super_mv v ON c.c_custkey = v.c_custkey;
```

Del mismo modo, puede reescribir la segunda consulta PartiQL. En el siguiente ejemplo, se utiliza un filtro en o_idx = 0.

```
SELECT c.c_name, v.o_totalprice
FROM customer_orders_lineitem c
JOIN super_mv v ON c.c_custkey = v.c_custkey
WHERE v.o_idx = 0;
```

En el comando CREATE MATERIALIZED VIEW, especifique c_custkey como clave de distribución y de ordenación para super_mv. Amazon Redshift realiza una unión de combinación eficiente, suponiendo que c_custkey también es la clave de distribución y la clave de ordenación de customer_orders_lineitem. Si ese no es el caso, puede especificar c_custkey como la clave de ordenación y de distribución de customer_orders_lineitem de la siguiente manera.

```
ALTER TABLE customer_orders_lineitem
ALTER DISTKEY c_custkey, ALTER SORTKEY (c_custkey);
```

Utilice la instrucción EXPLAIN para comprobar que Amazon Redshift realiza una unión de combinación en las consultas reescritas.

```
EXPLAIN
      SELECT c.c_name, v.o_orderstatus
      FROM customer_orders_lineitem c JOIN super_mv v ON c.c_custkey = v.c_custkey;

      QUERY PLAN

 ------------------------------------------------------------------------------------------
      XN Merge Join DS_DIST_NONE  (cost=0.00..34701.82 rows=1470776 width=27)
      Merge Cond: ("outer".c_custkey = "inner".c_custkey)
      ->  XN Seq Scan on mv_tbl__super_mv__0 derived_table2  (cost=0.00..14999.86
 rows=1499986 width=13)
      ->  XN Seq Scan on customer_orders_lineitem c  (cost=0.00..999.96 rows=99996
 width=30)
```

```
      (4 rows)
```

## Creación de columnas escalares de Amazon Redshift a partir de datos fragmentados

Los datos sin esquemas almacenados en SUPER pueden afectar el rendimiento de Amazon Redshift. Por ejemplo, los predicados de filtro o las condiciones de unión como análisis de rango restringido no pueden utilizar eficazmente los mapas de zona. Los usuarios y las herramientas de BI pueden utilizar las vistas materializadas como la presentación convencional de los datos y aumentar el rendimiento de las consultas analíticas.

La siguiente consulta analiza la vista materializada `super_mv` y filtra en `o_orderstatus`.

```
SELECT c.c_name, v.o_totalprice
FROM customer_orders_lineitem c
JOIN super_mv v ON c.c_custkey = v.c_custkey
WHERE v.o_orderstatus = 'F';
```

Inspeccione `stl_scan` para comprobar que Amazon Redshift no puede utilizar eficazmente los mapas de zona en el análisis de rango restringido a través de `o_orderstatus`.

```
SELECT slice, is_rrscan FROM stl_scan
WHERE query = pg_last_query_id() AND perm_table_name LIKE '%super_mv%';

 slice | is_rrscan
-------+-----------
     0 | f
     1 | f
     5 | f
     4 | f
     2 | f
     3 | f
(6 rows)
```

En el siguiente ejemplo, se adapta la vista materializada `super_mv` para crear columnas escalares a partir de los datos fragmentados. En este caso, Amazon Redshift convierte `o_orderstatus` de SUPER en VARCHAR. Además, especifique `o_orderstatus` como la clave de ordenación para `super_mv`.

```
CREATE MATERIALIZED VIEW super_mv distkey(c_custkey) sortkey(c_custkey, o_orderstatus)
 AS (
```

```
  SELECT c_custkey, o.o_orderstatus::VARCHAR AS o_orderstatus, o.o_totalprice, o_idx
  FROM customer_orders_lineitem c, c.c_orders o AT o_idx
);
```

Después de volver a ejecutar la consulta, compruebe que Amazon Redshift ahora pueda utilizar mapas de zona.

```
SELECT v.o_totalprice
FROM super_mv v
WHERE v.o_orderstatus = 'F';
```

Puede comprobar que el análisis de rango restringido ahora utiliza mapas de zona de la siguiente manera.

```
SELECT slice, is_rrscan FROM stl_scan
WHERE query = pg_last_query_id() AND perm_table_name LIKE '%super_mv%';

 slice | is_rrscan
-------+-----------
     0 | t
     1 | t
     2 | t
     3 | t
     4 | t
     5 | t
(6 rows)
```

## Limitaciones de uso del tipo de datos SUPER con vistas materializadas

Con Amazon Redshift, puede crear vistas materializadas con el tipo de datos SUPER para calcular previamente consultas complejas y mejorar el rendimiento de las consultas. El tipo de datos SUPER le permite almacenar el conjunto de resultados de una consulta como una estructura de datos plana, lo que permite un acceso y un procesamiento más rápidos.

En las siguientes secciones, se proporciona información detallada sobre las limitaciones y las prácticas recomendadas para utilizar el tipo de datos SUPER con vistas materializadas en Amazon Redshift.

Las vistas materializadas en Amazon Redshift no tienen ninguna limitación específica con respecto a PartiQL o SUPER.

Para obtener más información acerca de las limitaciones generales de SQL que se presentan cuando se crean vistas materializadas, consulte Limitaciones.

Para obtener más información acerca de las limitaciones generales de SQL que se presentan en las actualizaciones progresivas de vistas materializadas, consulte Limitaciones para la actualización incremental.

# Machine learning

El machine learning de Amazon Redshift (Amazon Redshift ML) es un servicio robusto basado en la nube que permite a los analistas y los científicos de datos con todos los niveles de habilidades usar tecnología de machine learning con más facilidad. Amazon Redshift ML utiliza un modelo para generar resultados. Puede utilizar modelos de las siguientes maneras:

- Puede proporcionar los datos que desea para entrenar un modelo y los metadatos asociados a las entradas de datos a Amazon Redshift. A continuación, Amazon Redshift ML crea modelos en Amazon SageMaker AI capaces de detectar patrones en los datos de entrada. Al utilizar sus propios datos para el modelo, puede utilizar Amazon Redshift ML para identificar tendencias en los datos, como la predicción de la pérdida, el valor de por vida del cliente o la predicción de los ingresos. Puede utilizar estos modelos para generar predicciones para los nuevos datos de entrada sin incurrir en costos adicionales.
- Puede utilizar uno de los modelos fundacionales (FM) proporcionados por Amazon Bedrock, como Claude o Amazon Titan. Con Amazon Bedrock, puede combinar la potencia de los modelos de lenguaje grande (LLM) con los datos de análisis en Amazon Redshift en unos pocos pasos. Si utiliza un modelo de lenguaje grande (LLM) externo, puede utilizar Amazon Redshift para realizar el procesamiento de lenguaje natural (NLP) en los datos. Puede utilizar el NLP para aplicaciones como la generación de texto, el análisis de opiniones o la traducción. Para obtener información acerca del uso de Amazon Bedrock con Amazon Redshift, consulte Integración de Amazon Redshift ML con Amazon Bedrock.

> **Note**
>
> Desactivación del uso de los datos para mejorar el servicio
> Si utiliza los modelos de Amazon Bedrock, le recomendamos que lea las políticas de AWS sobre cómo el servicio de Amazon Bedrock gestiona los datos. Debe determinar si necesita utilizar una política de exclusión voluntaria para evitar que el servicio utilice los datos para mejorar el modelo o el servicio, en caso de que Amazon Bedrock implemente dicha funcionalidad en el futuro. Para garantizar que el servicio no utilice los datos para tales fines, utilice la política general de exclusión de AWS.
> Para obtener más información, consulte los siguientes temas:
>
> - Políticas de exclusión de servicios de IA

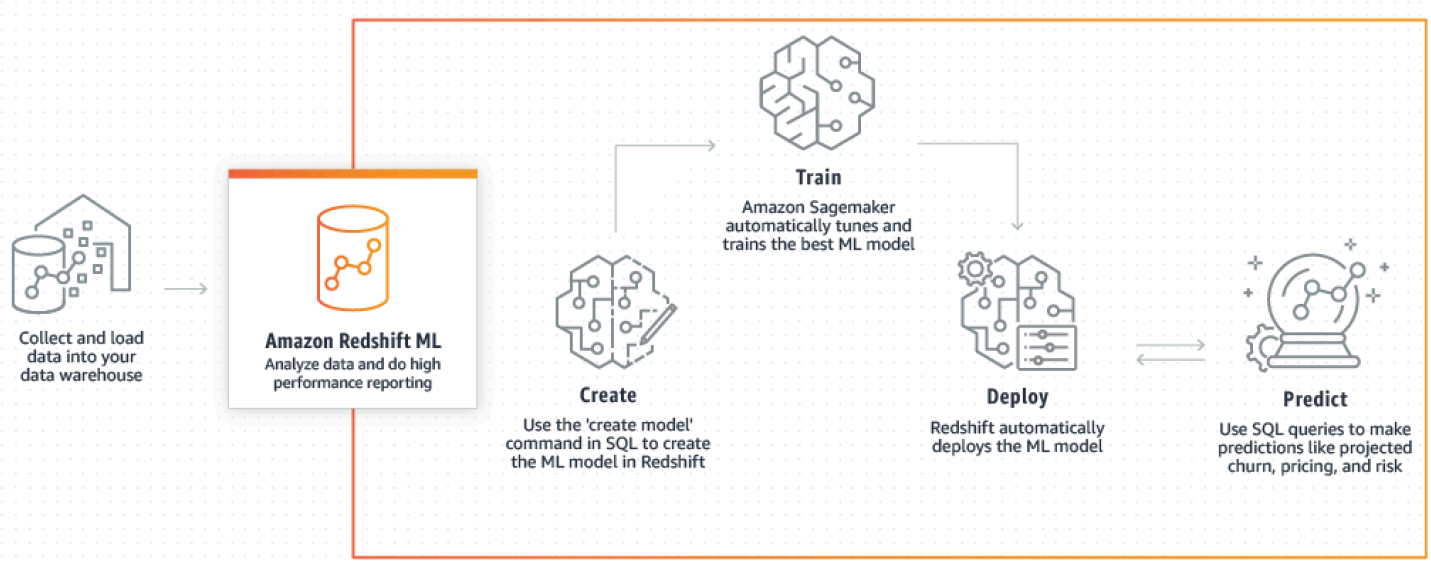
- Preguntas frecuentes de Amazon Bedrock

> **Note**
>
> Los LLM pueden generar información imprecisa o incompleta. Recomendamos verificar la información que producen los LLM para garantizar que sea precisa y completa.

Funcionamiento de Amazon Redshift ML con Amazon SageMaker AI

Amazon Redshift funciona con el Piloto automático de Amazon SageMaker AI para obtener automáticamente el modelo más adecuado y hacer que la función de predicción esté disponible en Amazon Redshift.

En el siguiente diagrama, se ilustra cómo funciona Amazon Redshift ML.




El flujo de trabajo general es el siguiente:

1. Amazon Redshift exporta los datos de formación a Amazon S3.
2. El Piloto automático de Amazon SageMaker AI procesa previamente los datos de entrenamiento. El procesamiento previo cumple funciones esenciales, tales como la imputación de los valores faltantes. Reconoce que ciertas columnas son categóricas (como el código postal), les da el formato adecuado para la formación y realiza muchas otras tareas. Elegir los mejores preprocesadores que se aplicarán en el conjunto de datos de entrenamiento es un problema en sí mismo, por lo que el Piloto automático de Amazon SageMaker AI automatiza su solución.

3. El Piloto automático de Amazon SageMaker AI encuentra el algoritmo y los hiperparámetros del algoritmo que ofrecen el modelo que tiene las predicciones más precisas.
4. Amazon Redshift registra la función de predicción como una función SQL en el clúster de Amazon Redshift.
5. Cuando se ejecutan instrucciones CREATE MODEL, Amazon Redshift utiliza Amazon SageMaker AI para el entrenamiento. Por lo tanto, la formación de su modelo conlleva un costo adicional. Se trata de una partida independiente específica de Amazon SageMaker AI en su factura de AWS. También se paga por el almacenamiento utilizado en Amazon S3 para guardar los datos de formación. No se cobra la inferencia mediante modelos creados con CREATE MODEL que puede compilar y ejecutar en su clúster de Redshift. No se aplican cargos adicionales a Amazon Redshift por utilizar Amazon Redshift ML.

Temas


- Información general sobre machine learning
- Machine learning para principiantes y expertos
- Costos de utilizar Amazon Redshift ML
- Getting started with Amazon Redshift ML (Introducción a Amazon Redshift ML)
- Tutoriales de Amazon Redshift ML
- Integración de Amazon Redshift ML con Amazon Bedrock


# Información general sobre machine learning

Con Amazon Redshift, puede aprovechar las capacidades de machine learning para obtener información valiosa de los datos. Esta información general de machine learning (ML) le muestra cómo explorar, visualizar y preparar los datos para el entrenamiento del modelo y la implementación de ML. Las siguientes secciones lo guiarán a través del proceso de uso de Amazon Redshift ML para desbloquear el potencial de los datos mediante machine learning.

A través de Amazon Redshift ML, puede formar modelos de machine learning con instrucciones SQL e invocarlos en consultas SQL para generar predicciones.

Si desea aprender a utilizar Amazon Redshift ML, puede ver el video que aparece a continuación: Amazon Redshift ML.

Para obtener información sobre los requisitos previos necesarios para configurar el clúster de Redshift o el grupo de trabajo sin servidor, los permisos y la propiedad para utilizar Amazon Redshift

ML, lea las siguientes secciones. En estas secciones, también se describe el funcionamiento sencillo de la formación y las predicciones en Amazon Redshift ML.

## Solución de problemas con machine learning

Un modelo de machine learning genera predicciones mediante la búsqueda de patrones en sus datos de formación y la posterior aplicación de dichos patrones a los nuevos datos. En machine learning, se forman estos modelos a través del aprendizaje de los patrones que mejor explican los datos. A continuación, se utilizan los modelos para generar predicciones (también denominadas inferencias) a partir de los nuevos datos. Por lo general, el machine learning es un proceso iterativo en el que se puede seguir mejorando la precisión de las predicciones a través de la modificación de los parámetros y la mejora de los datos de formación. Si se producen cambios en los datos, se vuelven a formar los nuevos modelos a partir del conjunto de datos nuevo.

Existen diferentes enfoques fundamentales en torno al machine learning para abordar los diversos objetivos empresariales.

### Aprendizaje supervisado en Amazon Redshift ML

Amazon Redshift admite el aprendizaje supervisado, que es el enfoque más común para el análisis empresarial avanzado. El aprendizaje supervisado constituye el enfoque de machine learning preferido cuando se dispone de un conjunto de datos establecido y de un conocimiento del modo en que los datos de entrada específicos predicen diversos resultados empresariales. Este tipo de resultado a veces se denomina etiqueta. En concreto, su conjunto de datos es una tabla con atributos que comprenden características (entradas) y objetivos (salidas). Por ejemplo, supongamos que dispone de una tabla que proporciona la edad y el código postal de los clientes anteriores y actuales. Supongamos que también dispone de un campo “activo” que se encuentra configurado como true para los clientes actuales y como false para quienes hayan suspendido su suscripción. El objetivo del machine learning supervisado consiste en detectar los patrones de edad y código postal que conducen a la deserción de los clientes, tal y como la representan los clientes cuyo objetivo es “False” (Falso). Puede utilizar este modelo para predecir qué clientes son propensos a darse de baja, como suspender su suscripción, para así ofrecer posibles incentivos para retenerlos.

Amazon Redshift admite el aprendizaje supervisado que incluye la regresión, la clasificación binaria y la clasificación multiclase. La regresión se refiere a la problemática de predecir valores continuos, como el gasto total de los clientes. La clasificación binaria se refiere a la problemática de predecir uno de dos resultados, como predecir si un cliente se dará de baja o no. La clasificación multiclase se refiere a la problemática de predecir uno de los muchos resultados, como predecir el elemento en el que podría estar interesado un cliente. Los analistas y los científicos de datos pueden utilizarla

para llevar a cabo un aprendizaje supervisado que les permita abordar una serie de problemáticas que van desde los pronósticos, la personalización o la predicción de la deserción de clientes. Asimismo, se puede recurrir al aprendizaje supervisado para resolver distintos problemas, como la predicción de las ventas que se concretarán, la predicción de los ingresos, la detección de fraudes y la predicción del valor de la vida útil del cliente.

## Aprendizaje sin supervisar en Amazon Redshift ML

El aprendizaje sin supervisar utiliza algoritmos de machine learning para analizar y agrupar datos de formación sin etiquetar. Los algoritmos descubren patrones o agrupaciones ocultos. El objetivo es modelar la estructura o distribución subyacente de los datos para obtener más información sobre los datos.

Amazon Redshift admite el algoritmo de clústeres K-Means para resolver un problema de aprendizaje sin supervisar. Este algoritmo resuelve problemas en clústeres en los que desea descubrir agrupaciones en los datos. El algoritmo K-Means intenta encontrar agrupaciones discretas dentro de los datos. Los datos sin clasificar se agrupan y dividen en función de sus similitudes y diferencias. Mediante la agrupación, el algoritmo K-Means determina iterativamente los mejores centroides y asigna cada miembro al centroide más cercano. Los miembros más cercanos al mismo centroide pertenecen al mismo grupo. Los miembros de un grupo son lo más parecido posible a otros miembros del mismo grupo y lo más diferente posible de los miembros de otros grupos. Por ejemplo, el algoritmo de clústeres K-Means se puede utilizar para clasificar ciudades afectadas por una pandemia o para clasificar ciudades en función de la popularidad de los productos de consumo.

Al utilizar el algoritmo K-Means, especifica una entrada k que especifica el número de clústeres para buscar en los datos. La salida de este algoritmo es un conjunto de centroides k. Cada punto de datos pertenece a uno de los clústeres k más cercanos a él. Cada clúster se describe mediante su centroide. El centroide se puede considerar como el promedio multidimensional del clúster. El algoritmo K-Means compara las distancias para ver qué tan diferentes son los clústeres entre sí. Generalmente, una distancia mayor indica una mayor diferencia entre los clústeres.

El procesamiento previo de los datos es importante para K-Means, ya que garantiza que las características del modelo se mantengan a la misma escala y produzcan resultados confiables. Amazon Redshift admite algunos procesadores previos de K-Means para la instrucción CREATE MODEL, como StandardScaler, MinMax y NumericPassthrough. Si no desea aplicar ningún procesamiento previo para K-Means, elija NumericPassthrough explícitamente como transformador. Para obtener más información sobre los parámetros de K-Means, consulte CREATE MODEL con parámetros K-MEANS.

Si desea aprender a realizar una formación sin supervisión con agrupación en clústeres de K-means, puede ver el siguiente video: Formación sin supervisión con agrupación de clústeres de K-means.

## Términos y conceptos de Amazon Redshift ML

Los siguientes términos se utilizan para describir algunos conceptos de Amazon Redshift ML:

- El machine learning en Amazon Redshift forma un modelo con un comando SQL. Amazon Redshift ML y Amazon SageMaker AI administran la totalidad de los procesos de conversión de datos, los permisos, el uso de recursos y la detección del modelo adecuado.
- La formación es la fase en la que Amazon Redshift crea un modelo de machine learning mediante la ejecución de un subconjunto específico de datos en el modelo. Amazon Redshift AI inicializa automáticamente un trabajo de entrenamiento en Amazon SageMaker AI y genera un modelo.
- La predicción (también denominada inferencia) representa el uso del modelo en las consultas SQL de Amazon Redshift para predecir resultados. En el momento de la inferencia, Amazon Redshift utiliza una función de predicción basada en un modelo que forma parte de una consulta más amplia destinada a generar predicciones. Las predicciones se calculan a nivel local, en el clúster de Redshift, lo que permite proporcionar un alto nivel de rendimiento, una baja latencia y ningún costo adicional.
- Con traiga su propio modelo (BYOM), puede utilizar un modelo entrenado fuera de Amazon Redshift con Amazon SageMaker AI para la inferencia local dentro de la base de datos de Amazon Redshift. Amazon Redshift ML admite el uso de BYOM en la inferencia local.
- La inferencia local se utiliza si los modelos se entrenan previamente en Amazon SageMaker AI, se compilan con Amazon SageMaker AI Neo y se localizan en Amazon Redshift ML. Si desea importar a Amazon Redshift modelos compatibles con la inferencia local, utilice el comando CREATE MODEL. Amazon Redshift importa los modelos ya entrenados de SageMaker AI llamando a Amazon SageMaker AI Neo. Allí, se compila el modelo y se importa el modelo compilado a Amazon Redshift. Utilice la inferencia local para obtener mayor velocidad y reducir los costos.
- La inferencia remota se utiliza cuando Amazon Redshift invoca un punto de conexión del modelo que se ha implementado en SageMaker AI. La inferencia remota proporciona la flexibilidad necesaria para invocar todo tipo de modelos personalizados y modelos de aprendizaje profundo; por ejemplo, los modelos de TensorFlow que se hayan creado e implementado en Amazon SageMaker AI.

Además, cabe destacar lo siguiente:

- Amazon SageMaker AI es un servicio de machine learning completamente administrado. Con Amazon SageMaker AI, los desarrolladores y los científicos de datos pueden crear, entrenar e implementar fácilmente y de manera directa los modelos en un entorno alojado listo para la producción. Para obtener más información acerca de Amazon SageMaker, consulte What is Amazon SageMaker AI en la Guía para desarrolladores de Amazon SageMaker AI.
- El Piloto automático de Amazon SageMaker AI es un conjunto de características que entrena y ajusta automáticamente los modelos de machine learning más adecuados para la clasificación o la regresión, en función de sus datos. Mantiene pleno control y visibilidad. El Piloto automático de Amazon SageMaker AI admite datos de entrada en formato tabular. El Piloto automático de Amazon SageMaker AI ofrece una limpieza y preprocesamiento automático de los datos, selección automática de algoritmos para la regresión lineal, clasificación binaria y clasificación multiclase. También admite la optimización automática de hiperparámetros (HPO), la formación distribuida, la instancia automática y la selección del tamaño del clúster. Para obtener más información acerca del Piloto automático de Amazon SageMaker AI, consulte Automate model development with Amazon SageMaker AI Autopilot en la Guía para desarrolladores de Amazon SageMaker AI.
- Amazon Bedrock es un servicio totalmente administrado que ofrece una selección de modelos fundacionales (FM) de alto rendimiento de las principales empresas de IA, como AI21 Labs, Anthropic, Cohere, Meta, Mistral AI, Stability AI y Amazon, a través de una sola API, junto con un amplio conjunto de capacidades necesarias para crear aplicaciones de IA generativa.

## Machine learning para principiantes y expertos

Con Amazon Redshift, puede aprovechar las capacidades de machine learning (ML) para obtener información a partir de los datos, si es un principiante o si es un experto en ML. Machine learning es una característica de Amazon Redshift que le permite crear, entrenar e implementar modelos de ML mediante comandos SQL, sin necesidad de una amplia experiencia en ML ni de una compleja ingeniería de datos.

Las siguientes secciones lo guiarán a través del proceso de uso de machine learning, lo que le permitirá desbloquear el potencial completo de los datos con Amazon Redshift.

Amazon Redshift ML le permite formar modelos con un solo comando SQL CREATE MODEL. El comando CREATE MODEL crea un modelo que Amazon Redshift utiliza para generar predicciones basadas en el modelo con construcciones SQL conocidas.

Amazon Redshift ML resulta de gran utilidad cuando no se tiene experiencia en lo que respecta a machine learning, herramientas, lenguajes, algoritmos y API. Gracias a Amazon Redshift ML, no se

ve obligado a realizar las tareas complejas y no diferenciadas que se requieren para la integración con un servicio de machine learning externo. Amazon Redshift le permite ahorrar tiempo para formatear y mover datos; administrar controles de permisos; o crear integraciones, flujos de trabajo y scripts personalizados. Se pueden utilizar fácilmente algoritmos de machine learning populares y simplificar las necesidades de formación que requieren una iteración frecuente desde la formación hasta la predicción. Amazon Redshift detecta de forma automática el mejor algoritmo y ajusta el modelo más adecuado para su problema. Puede generar predicciones desde el clúster de Amazon Redshift sin necesidad de trasladar los datos fuera de este ni de interconectarlos con otro servicio ni pagar por su uso.

Amazon Redshift ML ayuda a analistas de datos y científicos de datos a utilizar machine learning. También permite que los expertos en machine learning utilicen sus conocimientos para guiar la instrucción CREATE MODEL y utilizar únicamente los aspectos que especifican. De este modo, se puede acelerar el tiempo que CREATE MODEL necesita a la hora de buscar el mejor candidato, mejorar la precisión del modelo o ambos.

La instrucción CREATE MODEL ofrece flexibilidad en la manera en que puede especificar los parámetros para el trabajo de formación. Con esta flexibilidad, los principiantes o expertos en machine learning pueden elegir sus procesadores previos, algoritmos, tipos de problemas o hiperparámetros preferidos. Por ejemplo, un usuario interesado en la deserción de los clientes podría especificar para la instrucción CREATE MODEL que el tipo de problema corresponde a una clasificación binaria, la cual funciona bien para los casos de deserción de clientes. A continuación, la instrucción CREATE MODEL acota la búsqueda para encontrar el modelo más adecuado entre los modelos de clasificación binaria. Incluso con la elección del tipo de problema por parte del usuario, aún existen numerosas opciones con las que puede funcionar la instrucción CREATE MODEL. Por ejemplo, CREATE MODEL detecta y aplica las transformaciones de preprocesamiento más adecuadas y determina cuál es la mejor configuración de los hiperparámetros.

Amazon Redshift ML facilita el entrenamiento mediante la búsqueda automática del modelo más adecuado utilizando el Piloto automático de Amazon SageMaker AI. El Piloto automático de Amazon SageMaker AI entrena y ajusta automáticamente en segundo plano el modelo de machine learning más adecuado en función de los datos suministrados. A continuación, Amazon SageMaker AI Neo compila el modelo de entrenamiento y hace que esté disponible para la predicción en su clúster de Redshift. Cuando se ejecuta una consulta de inferencia de machine learning con un modelo formado, la consulta puede utilizar las capacidades de procesamiento masivo en paralelo de Amazon Redshift. Al mismo tiempo, la consulta puede utilizar predicciones basadas en machine learning.

- Los principiantes en machine learning, con conocimientos generales sobre los diferentes aspectos de esta tecnología, como procesadores previos, algoritmos e hiperparámetros, utilizan la instrucción CREATE MODEL únicamente para aquellos aspectos que especifican. De este modo, se puede acortar el tiempo que CREATE MODEL necesita a la hora de buscar el mejor candidato o mejorar la precisión del modelo. Además, se puede aumentar el valor empresarial de las predicciones mediante la incorporación de conocimientos adicionales sobre la materia, como el tipo de problema o el objetivo. Por ejemplo, en un escenario de deserción de clientes, si el resultado “el cliente no está activo” es poco frecuente, se suele preferir el objetivo F1 en lugar del objetivo de precisión. Como los modelos de alta precisión tal vez predicen que “el cliente está activo” en todo momento, esto se traduce en una alta precisión, pero escaso valor empresarial. Para obtener más información acerca de los objetivos F1, consulte AutoMLJobObjective en la Referencia de la API de Amazon SageMaker AI.

  Para obtener más información sobre las opciones básicas de la instrucción CREATE MODEL, consulte CREATE MODEL simple.
- Como profesional avanzado de machine learning, puede especificar el tipo de problema y los preprocesadores para ciertas características (aunque no todas). Luego, CREATE MODEL sigue las sugerencias sobre los aspectos especificados. Al mismo tiempo, CREATE MODEL continúa con la detección de los mejores procesadores previos para las características restantes y los mejores hiperparámetros. Para obtener más información sobre cómo restringir uno o más aspectos de la canalización de formación, consulte CREATE MODEL con guía para el usuario.
- Como experto en machine learning, puede tomar el control total de la formación y el ajuste de los hiperparámetros. Entonces, la instrucción CREATE MODEL no intenta detectar los procesadores previos, los algoritmos y los hiperparámetros óptimos, ya que es usted quien determina todas las opciones. Para obtener más información sobre cómo utilizar CREATE MODEL con AUTO OFF, consulte CREATE para modelos XGBoost con AUTO OFF.
- Como ingeniero de datos, puede incorporar un modelo de XGBoost previamente entrenado en Amazon SageMaker AI e importarlo a Amazon Redshift para la inferencia local. Con traiga su propio modelo (BYOM), puede utilizar un modelo entrenado fuera de Amazon Redshift con Amazon SageMaker AI para la inferencia local dentro de la base de datos de Amazon Redshift. Amazon Redshift ML admite el uso de BYOM en la inferencia local o remota.

  Para obtener más información sobre cómo utilizar la instrucción CREATE MODEL para inferencia local o remota, consulte Bring your own model (BYOM): inferencia local.

Los usuarios de Amazon Redshift ML pueden elegir cualquiera de las siguientes opciones para formar e implementar su modelo:

- Para los tipos de problemas, consulte CREATE MODEL con guía para el usuario.
- Para los objetivos, consulte CREATE MODEL con guía para el usuario o CREATE para modelos XGBoost con AUTO OFF.
- Para los tipos de modelo, consulte CREATE para modelos XGBoost con AUTO OFF.
- Para los preprocesadores, consulte CREATE MODEL con guía para el usuario.
- Para los hiperparámetros, consulte CREATE para modelos XGBoost con AUTO OFF.
- Para bring your own model (BYOM), consulte Bring your own model (BYOM): inferencia local.

# Costos de utilizar Amazon Redshift ML

Con Amazon Redshift, puede aprovechar las capacidades de machine learning para obtener información a partir de los datos sin necesidad de contar con una amplia experiencia en ingeniería de datos o machine learning. En las siguientes secciones se describen los costos asociados al uso de Amazon Redshift ML, lo que le ayuda a planificar y optimizar los gastos y, al mismo tiempo, aprovechar esta potente integración de machine learning.

## Costos por el uso de Amazon Redshift ML con SageMaker AI

Amazon Redshift ML para SageMaker AI usa los recursos de clúster existentes para la predicción, de modo que pueda evitar cargos adicionales de Amazon Redshift. No hay ningún cargo adicional de Amazon Redshift por crear o utilizar un modelo. La predicción se realiza localmente en el clúster de Redshift, por lo que no tiene que pagar más, a menos que necesite redimensionar el clúster. Amazon Redshift ML utiliza Amazon SageMaker AI para entrenar su modelo, lo que conlleva un costo adicional.

Las funciones de predicción que se ejecutan dentro de su clúster de Amazon Redshift no conllevan ningún cargo adicional. La instrucción CREATE MODEL utiliza Amazon SageMaker AI y conlleva un costo adicional. El costo aumenta en función del número de celdas de los datos de formación. El número de celdas se obtiene de multiplicar el número de registros (en la consulta de formación o en los tiempos de la tabla) por el número de columnas. Por ejemplo, cuando una consulta SELECT de la instrucción CREATE MODEL crea 10 000 registros y 5 columnas, se crean 50 000 celdas.

En algunos casos, los datos de formación obtenidos con la consulta SELECT de CREATE MODEL superan el límite de MAX_CELLS que se proporcionó (o el límite predeterminado de 1 millón en

caso de no haberlo hecho). En estos casos, CREATE MODEL elige de manera aleatoria un número aproximado para MAX_CELLS (es decir, los registros del “número de columnas” del conjunto de datos de formación). CREATE MODEL realiza la formación utilizando estas tuplas elegidas de manera aleatoria. Las muestras aleatorias garantizan que el conjunto de datos de formación reducido no tenga ningún tipo de sesgo. Por lo tanto, mediante la configuración de MAX_CELLS, puede controlar los costos de formación.

Cuando se utiliza la instrucción CREATE MODEL, se pueden utilizar las opciones MAX_CELLS y MAX_RUNTIME para controlar los costos, el tiempo y la posible precisión del modelo.

MAX_RUNTIME especifica la cantidad máxima de tiempo que puede durar el entrenamiento en SageMaker AI si se utiliza la opción AUTO ON u OFF. Los trabajos de formación suelen completarse antes que lo establecido en MAX_RUNTIME, en función del tamaño del conjunto de datos. Una vez que se haya formado un modelo, Amazon Redshift realiza más trabajo en segundo plano para compilar e instalar los modelos en el clúster. Por lo tanto, CREATE MODEL puede tardar más en completarse que lo establecido en MAX_RUNTIME. No obstante, MAX_RUNTIME limita la cantidad de tiempo y computación utilizados en SageMaker AI para entrenar el modelo. Puede verificar el estado de su modelo en cualquier momento con la opción SHOW MODEL.

Cuando se ejecuta CREATE MODEL con AUTO ON, Amazon Redshift ML recurre al Piloto automático de Amazon SageMaker AI para explorar de forma automática e inteligente diferentes modelos (o candidatos) con el objetivo de encontrar el más adecuado. MAX_RUNTIME limita la cantidad de tiempo y de cómputo empleados. Si MAX_RUNTIME se establece con un valor demasiado bajo, es posible que no se disponga de tiempo suficiente para explorar ni siquiera un candidato. Si ve el error “Autopilot candidate has no models” (El candidato de Autopilot no tiene modelos), vuelva a ejecutar CREATE MODEL con un valor para MAX_RUNTIME más grande. Para obtener más información sobre este parámetro, consulte MaxAutoMLJobRuntimeInSeconds en la Referencia de la API de Amazon SageMaker AI.

Cuando se ejecuta CREATE MODEL con AUTO OFF, MAX_RUNTIME representa un límite de tiempo de ejecución del trabajo de entrenamiento en SageMaker AI. Los trabajos de formación suelen completarse en menos tiempo en función del tamaño del conjunto de datos y de otros parámetros utilizados, como num_rounds en MODEL_TYPE XGBOOST.

También se pueden controlar los costos o acortar el tiempo de formación al especificar un valor menor para MAX_CELLS cuando se ejecuta CREATE MODEL. Una celda es una entrada en la base de datos. Cada fila corresponde a tantas celdas como columnas, cuyo ancho puede ser fijo o variable. MAX_CELLS limita el número de celdas y, por tanto, el número de ejemplos de formación utilizados a la hora de formar el modelo. De forma predeterminada, MAX_CELLS se configura en

1 millón de celdas. La reducción de MAX_CELLS disminuye el número de filas del resultado de la consulta SELECT en CREATE MODEL que Amazon Redshift exporta y envía a SageMaker AI para entrenar un modelo. Por lo tanto, la reducción de MAX_CELLS disminuye el tamaño del conjunto de datos utilizado para la formación de los modelos, tanto con AUTO ON como con AUTO OFF. Este enfoque permite reducir los costos y el tiempo de formación de los modelos. Para ver información sobre los tiempos facturación y entrenamiento de un trabajo de entrenamiento específico, elija Trabajos de entrenamiento en Amazon SageMaker AI.

El aumento de MAX_RUNTIME y MAX_CELLS suele mejorar la calidad del modelo, ya que permite a SageMaker AI ver más candidatos. De esta manera, SageMaker AI puede dedicar más tiempo entrenar a cada candidato y emplear más datos para entrenar mejores modelos. Si desea disponer de una iteración o exploración más rápida de su conjunto de datos, utilice MAX_RUNTIME y MAX_CELLS con valores más bajos. Si desea mejorar la precisión de los modelos, utilice MAX_RUNTIME y MAX_CELLS con valores más altos.

Para obtener más información sobre los costos asociados a los distintos números de celdas y los detalles de la prueba gratuita, consulte Precios de Amazon Redshift.

### Costos por el uso de Amazon Redshift ML con Amazon Bedrock

El uso de Amazon Redshift ML con Amazon Bedrock genera costos adicionales. Para obtener más información, consulte Precios de Amazon Bedrock.

## Getting started with Amazon Redshift ML (Introducción a Amazon Redshift ML)

Amazon Redshift ML permite a los usuarios de SQL crear, formar e implementar con más facilidad modelos de machine learning mediante comandos SQL conocidos. Con Amazon Redshift ML, puede utilizar los datos en el clúster de Redshift para entrenar modelos con Amazon SageMaker AI. Luego, los modelos se localizan y se pueden generar predicciones en una base de datos de Amazon Redshift. Amazon Redshift ML admite actualmente los algoritmos de machine learning: XGBoost (AUTO ON y OFF) y perceptron multicapa (AUTO ON), K-Means (AUTO OFF) y Linear Learner.

Temas

- Agrupe y ajuste la configuración para la administración de Amazon Redshift ML
- Uso de la explicabilidad del modelo con Amazon Redshift ML

- Métricas de probabilidad de Amazon Redshift ML

# Agrupe y ajuste la configuración para la administración de Amazon Redshift ML

Antes de trabajar con Amazon Redshift ML, complete la configuración del clúster y establezca los permisos para utilizar este servicio.

## Configuración del clúster para utilizar Amazon Redshift ML

Antes de trabajar con Amazon Redshift ML, complete los siguientes requisitos previos.

Como administrador de Amazon Redshift, complete la siguiente configuración única para utilizar clústeres aprovisionados de Amazon Redshift. Para utilizar Amazon Redshift ML con Amazon Redshift sin servidor, consulte Introducción a Amazon Redshift sin servidor.

Para realizar la configuración de clúster por única ver para Amazon Redshift ML

1. Cree un clúster de Redshift con la AWS Management Console o la AWS Command Line Interface (AWS CLI). Asegúrese de adjuntar la política de AWS Identity and Access Management (IAM) cuando crea el clúster. Para obtener más información sobre los permisos necesarios para utilizar Amazon Redshift ML con Amazon SageMaker AI, consulte Permisos necesarios para usar machine learning (ML) de Amazon Redshift
2. Cree el rol de IAM necesario para utilizar Amazon Redshift ML de una de las siguientes formas:

   - Para usar SageMaker AI con Amazon Redshift ML, cree un rol de IAM con las políticas `AmazonS3FullAccess` y `AmazonSageMakerFullAccess`. Si planea crear también modelos de previsión, adjunte también la política `AmazonForecastFullAccess` al rol.
   - Para usar Amazon Bedrock con Amazon Redshift ML, cree un rol de IAM con las políticas `AmazonS3FullAccess` y `AmazonBedrockFullAccess`.
   - Le recomendamos que cree un rol de IAM a través de la consola de Amazon Redshift que tiene la política de `AmazonRedshiftAllCommandsFullAccess` con permisos para ejecutar comandos SQL, como CREATE MODEL. Amazon Redshift utiliza un mecanismo fluido basado en la API para crear roles de IAM mediante programación en su Cuenta de AWS a su nombre. Amazon Redshift adjunta automáticamente políticas administradas existentes de AWS al rol de IAM. Este enfoque significa que puede permanecer dentro de la consola de Amazon Redshift y no tiene que cambiar a la consola de IAM para crear roles. Para obtener

más información, consulte Creación de un rol de IAM como predeterminado para Amazon Redshift.

Cuando se crea un rol de IAM como predeterminado para su clúster, incluya `redshift` como parte del nombre del recurso o utilice una etiqueta específica de RedShift para etiquetar esos recursos.

Si su clúster tiene el enrutamiento mejorado de Amazon VPC activado, puede utilizar un rol de IAM creado a través de la consola de Amazon Redshift. Este rol de IAM tiene adjunta la política de `AmazonRedshiftAllCommandsFullAccess` y agrega los siguientes permisos a la política. Estos permisos adicionales permiten que Amazon Redshift cree y elimine una interfaz de red elástica (ENI) de su cuenta y la adjunte a tareas de compilación que se ejecutan en Amazon EC2 o Amazon ECS. De este modo, permite que se acceda a los objetos en los bucketS de Amazon S3 solo desde una nube virtual privada (VPC) con acceso a la Internet bloqueado.

```
{
    "Effect": "Allow",
    "Action": [
        "ec2:DescribeVpcEndpoints",
        "ec2:DescribeDhcpOptions",
        "ec2:DescribeVpcs",
        "ec2:DescribeSubnets",
        "ec2:DescribeSecurityGroups",
        "ec2:DescribeNetworkInterfaces",
        "ec2:DeleteNetworkInterfacePermission",
        "ec2:DeleteNetworkInterface",
        "ec2:CreateNetworkInterfacePermission",
        "ec2:CreateNetworkInterface",
        "ec2:ModifyNetworkInterfaceAttribute"
    ],
    "Resource": "*"
}
```

Para usar los modelos fundacionales de Amazon Bedrock, agregue la siguiente sección:

```
// Required section if you use Bedrock models.
{
   "Effect": "Allow",
   "Action": "bedrock:InvokeModel",
   "Resource": [
```

```
        "arn:aws:bedrock:<region>::foundation-model/*"
    ]
}
```

- Si desea crear un rol de IAM que tenga una política más restrictiva, puede emplear la siguiente política. Asimismo, puede modificarla para que se adapte a sus necesidades.

  El bucket `redshift-downloads/redshift-ml/` de Amazon S3 es el lugar donde se almacenan los datos de muestra utilizados en otros pasos y ejemplos. Puede quitarlo si no necesita cargar datos de Amazon S3. O bien, puede reemplazarlo con otros buckets de Amazon S3 que utilice para cargar datos en Amazon Redshift.

  Los valores *`your-account-id`*, *`your-role`* y `amzn-s3-demo-bucket` son los que se especifican como parte del comando CREATE MODEL.

  (Opcional) Utilice la sección de claves de AWS KMS que corresponde a la política de muestra si especifica una clave de AWS KMS mientras utiliza Amazon Redshift ML. El valor *`your-kms-key`* es la clave que se utiliza como parte del comando CREATE MODEL.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "cloudwatch:PutMetricData",
                "ecr:BatchCheckLayerAvailability",
                "ecr:BatchGetImage",
                "ecr:GetAuthorizationToken",
                "ecr:GetDownloadUrlForLayer",
                "logs:CreateLogGroup",
                "logs:CreateLogStream",
                "logs:DescribeLogStreams",
                "logs:PutLogEvents",
                "sagemaker:*Job*",
                "sagemaker:AddTags",
                "sagemaker:CreateModel",
                "sagemaker:CreateEndpoint",
                "sagemaker:CreateEndpointConfig",
                "sagemaker:DeleteEndpoint",
```

```
                "sagemaker:DeleteEndpointConfig,
                "sagemaker:DeleteModel"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:PassRole",
                "s3:AbortMultipartUpload",
                "s3:GetObject",
                "s3:DeleteObject",
                "s3:PutObject"
            ],
            "Resource": [
                "arn:aws:iam::<your-account-id>:role/<your-role>",
                "arn:aws:s3:::amzn-s3-demo-bucket/*",
                "arn:aws:s3:::redshift-downloads/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:GetBucketLocation",
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::amzn-s3-demo-bucket,
                "arn:aws:s3:::redshift-downloads"
            ]
        }
        // Optional section needed if you use AWS KMS keys.
        ,{
            "Effect": "Allow",
            "Action": [
                "kms:CreateGrant",
                "kms:Decrypt",
                "kms:DescribeKey",
                "kms:Encrypt",
                "kms:GenerateDataKey*"
             ],
             "Resource": [
                "arn:aws:kms:<your-region>:<your-account-id>:key/<your-kms-key>"
             ]
```

```
        }
    ]
}
```

3. Para permitir que Amazon Redshift y SageMaker AI asuman el rol para interactuar con otros servicios, añada la siguiente política de confianza al rol de IAM.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": [
          "redshift.amazonaws.com",
          "sagemaker.amazonaws.com",
          "forecast.amazonaws.com"
        ]
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

4. (Opcional) Cree un bucket de Amazon S3 y una clave AWS KMS. Amazon Redshift los utiliza para almacenar los datos de entrenamiento enviados a Amazon SageMaker AI y recibir el modelo entrenado de Amazon SageMaker AI.
5. (Opcional) Cree diferentes combinaciones de roles de IAM y buckets de Amazon S3 que permitan controlar el acceso a diferentes grupos de usuarios.
6. (Opcional) Cuando active el enrutamiento de VPC para el clúster de Redshift, cree lo siguiente:

   - Un punto de conexión de Amazon S3 o la VPC en la que se encuentra el clúster de Redshift
   - Un punto de conexión de IA de SageMaker o un punto de conexión de VPC de Amazon Bedrock

   Esto posibilita que el tráfico pase por la VPC entre servicios durante el proceso CREATE MODEL. Para obtener más información sobre el enrutamiento de la VPC, consulte Enrutamiento de VPC mejorado en Amazon Redshift.

Para obtener más información sobre los permisos necesarios para especificar una VPC privada para el trabajo de ajuste de hiperparámetros, consulte Permisos necesarios para usar el machine learning de Amazon Redshift con Amazon SageMaker AI.

> **Note**
>
> Las llamadas de inferencia realizadas a modelos remotos de IA de SageMaker no pasan por LA VPC.

Para obtener información sobre cómo utilizar la instrucción CREATE MODEL con objeto de comenzar a crear modelos para diferentes casos de uso, consulte CREATE MODEL.

## Administración de permisos y propiedad

Al igual que con otros objetos de base de datos, como tablas o funciones, Amazon Redshift vincula la creación y el uso de modelos de ML para acceder a los mecanismos de control. Para crear un modelo que ejecute funciones de predicción, se dispone de distintos permisos.

En los siguientes ejemplos, se utilizan dos grupos de usuarios, `retention_analyst_grp` (creador del modelo) y `marketing_analyst_grp` (usuario del modelo) para demostrar cómo Amazon Redshift administra el control de acceso. El analista de retención crea modelos de machine learning que el otro conjunto de usuarios puede emplear gracias a los permisos obtenidos.

Un superusuario puede utilizar el permiso GRANT USER o GROUP para crear modelos de machine learning con la siguiente instrucción.

```
GRANT CREATE MODEL TO GROUP retention_analyst_grp;
```

Los usuarios o los grupos que tengan este permiso pueden crear un modelo en cualquier esquema del clúster siempre que el usuario cuente con el permiso habitual CREATE en el SCHEMA. El modelo de machine learning forma parte de la jerarquía de esquemas del mismo modo que las tablas, las vistas, los procedimientos y las funciones definidas por el usuario.

En el supuesto de que ya exista un esquema `demo_ml`, conceda a los dos grupos de usuarios el permiso correspondiente al esquema de la siguiente manera.

```
GRANT CREATE, USAGE ON SCHEMA demo_ml TO GROUP retention_analyst_grp;
```

```
GRANT USAGE ON SCHEMA demo_ml TO GROUP marketing_analyst_grp;
```

Para permitir que otros usuarios utilicen la función de inferencia de machine learning, conceda el permiso EXECUTE. En el siguiente ejemplo, se utiliza el permiso EXECUTE para conceder al GRUPO marketing_analyst_grp el permiso de usar el modelo.

```
GRANT EXECUTE ON MODEL demo_ml.customer_churn_auto_model TO GROUP
 marketing_analyst_grp;
```

Utilice la instrucción REVOKE con CREATE MODEL y EXECUTE para revocar esos permisos que se habían otorgado a los usuarios o los grupos. Para obtener más información sobre comandos de control de permisos, consulte GRANT y REVOKE.

## Uso de la explicabilidad del modelo con Amazon Redshift ML

Con la explicabilidad del modelo en Amazon Redshift (ML), utiliza valores de importancia de características para ayudar a comprender cómo contribuye cada atributo de los datos de formación al resultado previsto.

La explicabilidad del modelo ayuda a mejorar los modelos de machine learning (ML) al explicar las predicciones que hacen los modelos. La explicabilidad del modelo ayuda a explicar cómo estos modelos hacen predicciones utilizando un enfoque de atribución de características.

Amazon Redshift ML incorpora la explicabilidad del modelo para proporcionar la funcionalidad de explicación del modelo a los usuarios de Amazon Redshift ML. Para obtener más información sobre el modelo de explicabilidad, consulte ¿Qué es la equidad y la explicabilidad del modelo en las predicciones de machine learning? en la Guía para desarrolladores de Amazon SageMaker AI.

La explicabilidad del modelo también monitorea las inferencias que hacen los modelos en producción para la desviación de la atribución de características. También proporciona herramientas para ayudarlo a generar informes de gobernanza de modelos que puede utilizar para informar a los equipos de riesgo y conformidad, así como a los reguladores externos.

Cuando especifica la opción AUTO ON o AUTO OFF al utilizar la instrucción CREATE MODEL, una vez finalizado el trabajo de entrenamiento del modelo, SageMaker AI crea la salida de la explicación. Puede utilizar la función EXPLAIN_MODEL para consultar el informe de explicabilidad en formato JSON. Para obtener más información, consulte Funciones de machine learning..

# Métricas de probabilidad de Amazon Redshift ML

En los problemas de aprendizaje supervisado, las etiquetas de clase son el resultado de predicciones que utilizan los datos de entrada. Por ejemplo, si utiliza un modelo para predecir si un cliente volvería a suscribirse a un servicio de streaming, las posibles etiquetas son “probable” e “improbable”. Redshift ML proporciona la capacidad de las métricas de probabilidad, que asignan una probabilidad a cada etiqueta para indicar su probabilidad. Esto le ayuda a tomar decisiones más informadas en función de los resultados previstos. En Amazon Redshift ML, las métricas de probabilidad están disponibles al crear modelos AUTO ON con un tipo de problema de clasificación binaria o clasificación multiclase. Si omite el parámetro AUTO ON, Redshift ML asume que el modelo debe tener AUTO ON.

## Creación del modelo

Al crear un modelo, Amazon Redshift detecta automáticamente el tipo de modelo y el tipo de problema. Si se trata de un problema de clasificación, Redshift crea automáticamente una segunda función de inferencia que puede utilizar para generar probabilidades en relación con cada etiqueta. El nombre de esta segunda función de inferencia es el nombre de la función de inferencia especificada seguido de la cadena `_probabilities`. Por ejemplo, si asigna a la función de inferencia el nombre `customer_churn_predict`, el nombre de la segunda función de inferencia es `customer_churn_predict_probabilities`. A continuación, puede consultar esta función para obtener las probabilidades de cada etiqueta.

```
CREATE MODEL customer_churn_model
FROM customer_activity
    PROBLEM_TYPE BINARY_CLASSIFICATION
TARGET churn
FUNCTION customer_churn_predict
IAM_ROLE {default}
AUTO ON
SETTINGS ( S3_BUCKET 'amzn-s3-demo-bucket'
```

## Obtención de probabilidades

Una vez que la función de probabilidad esté lista, la ejecución del comando devuelve un tipo SUPER que contiene matrices de las probabilidades devueltas y sus etiquetas asociadas. Por ejemplo, el resultado `"probabilities" : [0.7, 0.3], "labels" : ["False.", "True."]` significa que la etiqueta False tiene una probabilidad de 0,7 y la etiqueta True tiene una probabilidad de 0,3.

```
SELECT customer_churn_predict_probabilities(Account_length, Area_code,
            VMail_message, Day_mins, Day_calls, Day_charge,Eve_mins, Eve_calls,
            Eve_charge, Night_mins, Night_calls, Night_charge,Intl_mins, Intl_calls,
            Intl_charge, Cust_serv_calls)
FROM customer_activity;

customer_churn_predict_probabilities
 --------------------
 {"probabilities" : [0.7, 0.3], "labels" : ["False.", "True."]}
 {"probabilities" : [0.8, 0.2], "labels" : ["False.", "True."]}
 {"probabilities" : [0.75, 0.25], "labels" : ["True.", "False"]}
```

Las matrices de probabilidades y etiquetas siempre se ordenan por sus probabilidades en orden descendente. Puede escribir una consulta para que devuelva solo la etiqueta predicha con la probabilidad más alta mediante la anulación de la anidación de los resultados SUPER devueltos de la función de probabilidad.

```
SELECT prediction.labels[0], prediction.probabilities[0]
            FROM (SELECT customer_churn_predict_probabilities(Account_length,
 Area_code,
            VMail_message, Day_mins, Day_calls, Day_charge,Eve_mins, Eve_calls,
            Eve_charge, Night_mins, Night_calls, Night_charge,Intl_mins, Intl_calls,
            Intl_charge, Cust_serv_calls) AS prediction
FROM customer_activity);

  labels   | probabilities
-----------+--------------
 "False."  | 0.7
 "False."  | 0.8
 "True."   | 0.75
```

Para simplificar las consultas, puede almacenar los resultados de la función de predicción en una tabla.

```
CREATE TABLE churn_auto_predict_probabilities AS
             (SELECT customer_churn_predict_probabilities(Account_length, Area_code,
             VMail_message, Day_mins, Day_calls, Day_charge,Eve_mins, Eve_calls,
             Eve_charge, Night_mins, Night_calls, Night_charge,Intl_mins,
             Intl_calls, Intl_charge, Cust_serv_calls) AS prediction
FROM customer_activity);
```

Puede consultar la tabla con los resultados para devolver solo las predicciones que tengan una probabilidad superior a 0,7.

```
SELECT prediction.labels[0], prediction.probabilities[0]
FROM churn_auto_predict_probabilities
WHERE prediction.probabilities[0] > 0.7;

  labels   | probabilities
-----------+--------------
 "False."  | 0.8
 "True."   | 0.75
```

Mediante la notación de índice, puede obtener la probabilidad de una etiqueta específica. En el ejemplo siguiente se devuelven las probabilidades de todas las etiquetas `True.`.

```
SELECT label, index, p.prediction.probabilities[index]
FROM churn_auto_predict_probabilities p, p.prediction.labels AS label AT index
WHERE label='True.';

  label  | index | probabilities
---------+-------+---------------
 "True." |     0 | 0.3
 "True." |     0 | 0.2
 "True." |     0 | 0.75
```

En el siguiente ejemplo se devuelven todas las filas que tienen una etiqueta `True` con una probabilidad superior a 0,7, lo que indica que es probable que el cliente abandone.

```
SELECT prediction.labels[0], prediction.probabilities[0]
FROM churn_auto_predict_probabilities
WHERE prediction.probabilities[0] > 0.7 AND prediction.labels[0] = "True.";

labels     | probabilities
-----------+--------------
 "True."   | 0.75
```

# Tutoriales de Amazon Redshift ML

Puede utilizar Amazon Redshift ML para entrenar modelos de machine learning mediante instrucciones SQL y, a continuación, invocar los modelos en consultas SQL para realizar

predicciones. El machine learning en Amazon Redshift forma un modelo con un comando SQL. Amazon Redshift AI inicializa automáticamente un trabajo de entrenamiento en Amazon SageMaker AI y genera un modelo. Una vez creado el modelo, puede realizar predicciones en Amazon Redshift mediante la función de predicción del modelo.

Siga los pasos de estos tutoriales para obtener más información acerca de las características de Amazon Redshift ML:

- Tutorial: Creación de modelos de deserción de clientes: en este tutorial, utilizará Amazon Redshift ML para crear un modelo de deserción de clientes con el comando CREATE MODEL y ejecutará consultas de predicción para situaciones de usuario. A continuación, implementará consultas mediante la función SQL que genera el comando CREATE MODEL.
- Tutorial: Creación de modelos de agrupación en clústeres k-means: en este tutorial, utilizará Amazon Redshift ML para crear, entrenar e implementar un modelo de machine learning basado en el algoritmo k-means.
- Tutorial: Creación de modelos de clasificación multiclase: en este tutorial, utilizará Amazon Redshift ML para crear un modelo de machine learning que resuelva los problemas de clasificación multiclase. El algoritmo de clasificación multiclase clasifica los puntos de datos en una de tres o más clases. A continuación, implementará consultas mediante la función SQL que genera el comando CREATE MODEL.
- Tutorial: Creación de modelos XGBoost: en este tutorial, creará un modelo con datos de Amazon S3 y ejecutará consultas de predicción con el modelo mediante Amazon Redshift ML. El algoritmo XGBoost es una implementación optimizada del algoritmo de árboles aumentados.
- Tutorial: Creación de modelos de regresión: en este tutorial, utilizará Amazon Redshift ML para crear un modelo de regresión de machine learning y ejecutar consultas de predicción en el modelo. Los modelos de regresión le permiten predecir resultados numéricos, como el precio de una casa o cuántas personas utilizarán el servicio de alquiler de bicicletas de una ciudad.
- Tutorial: Creación de modelos de regresión con aprendizaje lineal: en este tutorial, creará un modelo de aprendizaje lineal con datos de Amazon S3 y ejecutará consultas de predicción con el modelo mediante Amazon Redshift ML. El algoritmo de aprendizaje lineal de SageMaker AI resuelve problemas de regresión o clasificación multiclase.
- Tutorial: Creación de modelos de clasificación multiclase con aprendizaje lineal: en este tutorial, creará un modelo de aprendizaje lineal con datos de Amazon S3 y, a continuación, ejecutará consultas de predicción con el modelo mediante Amazon Redshift ML. El algoritmo de aprendizaje lineal de SageMaker AI resuelve problemas de regresión o clasificación.

# Tutorial: Creación de modelos de deserción de clientes

En este tutorial, utilizará Amazon Redshift ML para crear un modelo de deserción de clientes con el comando CREATE MODEL y ejecutará consultas de predicción para situaciones de usuario. A continuación, implementará consultas mediante la función SQL que genera el comando CREATE MODEL.

Se puede utilizar un comando sencillo CREATE MODEL para exportar datos de formación, formar un modelo, importar ese modelo y preparar una función de predicción de Amazon Redshift. Utilice la instrucción CREATE MODEL para especificar los datos de formación en forma de tabla o con la instrucción SELECT.

Este ejemplo utiliza información histórica para construir un modelo de machine learning sobre las bajas de clientes de un operador de telefonía móvil. En primer lugar, SageMaker AI entrena el modelo de machine learning y, a continuación, lo prueba utilizando la información de perfil de un cliente arbitrario. Una vez validado el modelo, Amazon SageMaker AI implementa el modelo y la función de predicción en Amazon Redshift. Puede usar la función de predicción para predecir si un cliente va a darse de baja o no.

## Ejemplos de casos de uso

Puede resolver otros problemas de clasificación binaria con Amazon Redshift ML, como predecir si un cliente potencial de ventas se cerrará o no. También puede predecir si una transacción financiera es fraudulenta o no.

Tareas

- Requisitos previos
- Paso 1: Cargar los datos desde Amazon S3 en Amazon Redshift
- Paso 2: Crear el modelo de machine learning
- Paso 3: Realizar predicciones con el modelo

## Requisitos previos

Para completar este tutorial, necesita los siguientes requisitos previos:

- Debe configurar un clúster de Amazon Redshift para Amazon Redshift ML. Para ello, utilice la documentación relativa a la Agrupación y ajuste de la configuración para la administración de Amazon Redshift ML.

- El clúster de Amazon Redshift que se utiliza para crear el modelo y el bucket de Amazon S3 que se utiliza para preparar los datos de entrenamiento y almacenar los artefactos del modelo deberán estar en la misma región de AWS.
- Para descargar los comandos SQL y el conjunto de datos de muestra utilizados en esta documentación, realice una de las siguientes acciones:
  - Descargue los comandos SQL, el archivo de actividad del cliente y el archivo de Abalone.
  - A través de la AWS CLI para Amazon S3, ejecute el siguiente comando. Puede utilizar su propia ruta de destino.

```
aws s3 cp s3://redshift-downloads/redshift-ml/tutorial-scripts/redshift-ml-
tutorial.sql </target/path>
aws s3 cp s3://redshift-downloads/redshift-ml/customer_activity/
customer_activity.csv </target/path>
aws s3 cp s3://redshift-downloads/redshift-ml/abalone_xgb/abalone_xgb.csv </target/
path>
```

## Paso 1: Cargar los datos desde Amazon S3 en Amazon Redshift

Use el editor de consultas de Amazon Redshift v2 para editar y ejecutar consultas y visualizar los resultados.

Al ejecutar las consultas siguientes, se crea una tabla llamada `customer_activity` y se incorpora el conjunto de datos de ejemplo desde Amazon S3.

```
DROP TABLE IF EXISTS customer_activity;

CREATE TABLE customer_activity (
state varchar(2),
account_length int,
area_code int,
phone varchar(8),
intl_plan varchar(3),
vMail_plan varchar(3),
vMail_message int,
day_mins float,
day_calls int,
day_charge float,
total_charge float,
eve_mins float,
eve_calls int,
```

```
eve_charge float,
night_mins float,
night_calls int,
night_charge float,
intl_mins float,
intl_calls int,
intl_charge float,
cust_serv_calls int,
churn varchar(6),
record_date date
);

COPY customer_activity
FROM 's3://redshift-downloads/redshift-ml/customer_activity/'
REGION 'us-east-1' IAM_ROLE default
FORMAT AS CSV IGNOREHEADER 1;
```

## Paso 2: Crear el modelo de machine learning

La deserción es nuestro objetivo de entrada en este modelo. Todas las demás entradas del modelo son atributos que ayudan a crear una función para predecir la deserción.

En el siguiente ejemplo, se utiliza la operación CREATE MODEL para ofrecer un modelo que prediga si un cliente permanecerá activo. Para ello, se utilizan distintos datos, como edad del cliente, código postal, gasto y casos. En el siguiente ejemplo, reemplace amzn-s3-demo-bucket por el bucket de Amazon S3 propio.

```
CREATE MODEL customer_churn_auto_model
FROM
    (
      SELECT state,
             account_length,
             area_code,
             total_charge/account_length AS average_daily_spend,
             cust_serv_calls/account_length AS average_daily_cases,
             churn
      FROM customer_activity
      WHERE  record_date < '2020-01-01'
     )
TARGET churn FUNCTION ml_fn_customer_churn_auto
IAM_ROLE default SETTINGS (
  S3_BUCKET 'amzn-s3-demo-bucket'
```

```
);
```

La consulta SELECT del ejemplo anterior crea los datos de entrenamiento. La cláusula TARGET especifica cuál es la columna que corresponde a la etiqueta de machine learning que utiliza la operación CREATE MODEL para aprender a generar predicciones. La columna de destino "churn" (deserción) indica si el cliente todavía tiene una suscripción activa o si la ha suspendido. El campo S3_BUCKET es el nombre del bucket de Amazon S3 que ha creado anteriormente. El bucket de Amazon S3 se utiliza para compartir los datos y artefactos de entrenamiento entre Amazon Redshift y Amazon SageMaker AI. Las columnas restantes corresponden a las características que se utilizan para la predicción.

Para ver un resumen de la sintaxis y las características de un caso de uso básico del comando CREATE MODEL, consulte CREATE MODEL simple.

Añadir permisos para el cifrado del lado del servidor (opcional)

Amazon Redshift utiliza de forma predeterminada el Piloto automático de Amazon SageMaker AI para el entrenamiento. En concreto, Amazon Redshift exporta de forma segura los datos de entrenamiento al bucket de Amazon S3 que especifica el cliente. Si no se especifica un `KMS_KEY_ID`, los datos se cifran mediante el cifrado del servidor SSE-S3 de forma predeterminada.

Si cifra la entrada mediante el cifrado del lado del servidor con una clave administrada por AWS KMS (SSE-MMS), agregue los siguientes permisos:

```
{
    "Effect": "Allow",
    "Action": [
    "kms:Encrypt"
    "kms:Decrypt"
    ]
}
```

Para obtener más información acerca de los roles de Amazon SageMaker, consulte Amazon SageMaker AI roles en la Guía para desarrolladores de Amazon SageMaker AI.

Comprobar el estado del entrenamiento del modelo (opcional)

Puede utilizar el comando SHOW MODEL para saber cuándo está listo el modelo.

Utilice la siguiente operación para comprobar el estado del modelo.

```
SHOW MODEL customer_churn_auto_model;
```

A continuación se muestra un ejemplo de la salida de la operación anterior.

```
+--------------------------
+-----------------------------------------------------------------------------------------------------------
+
|           Key            |
                 Value
             |
+--------------------------
+-----------------------------------------------------------------------------------------------------------
+
|        Model Name        |
        customer_churn_auto_model
             |
|       Schema Name        |
                 public
             |
|          Owner           |
                 awsuser
             |
|      Creation Time       |
        Tue, 14.06.2022 17:15:52
             |
|       Model State        |
                TRAINING
             |
|                          |

             |
|      TRAINING DATA:      |

             |
|          Query           | SELECT STATE, ACCOUNT_LENGTH, AREA_CODE, TOTAL_CHARGE /
 ACCOUNT_LENGTH AS AVERAGE_DAILY_SPEND, CUST_SERV_CALLS / ACCOUNT_LENGTH AS
 AVERAGE_DAILY_CASES, CHURN |
|                          |
         FROM CUSTOMER_ACTIVITY
             |
|                          |
    WHERE RECORD_DATE < '2020-01-01'
             |
```

```
|      Target Column       |
                  CHURN
             |
|                          |

             |
|       PARAMETERS:        |

             |
|        Model Type        |
                  auto
             |
|       Problem Type       |

             |
|        Objective         |

             |
|     AutoML Job Name      |
     redshiftml-20220614171552640901
             |
|      Function Name       |
        ml_fn_customer_churn_auto
             |
|   Function Parameters    |                                                  state
 account_length area_code average_daily_spend average_daily_cases
                       |
| Function Parameter Types |
      varchar int4 int4 float8 int4
             |
|         IAM Role         |
          default-aws-iam-role
             |
|        S3 Bucket         |
             amzn-s3-demo-bucket
              |
|       Max Runtime        |
                  5400
             |
+--------------------------
+-----------------------------------------------------------------------------------------------------------
+
```

Una vez finalizado el entrenamiento del modelo, la variable `model_state` se convierte en `Model is Ready` y la función de predicción pasa a estar disponible.

## Paso 3: Realizar predicciones con el modelo

Puede usar instrucciones SQL para ver las predicciones realizadas por el modelo de predicción. En este ejemplo, la función de predicción creada por la operación CREATE MODEL se denomina `ml_fn_customer_churn_auto`. Los argumentos de entrada para la función de predicción corresponden a los tipos de características, como varchar para `state` y entero para `account_length`. La salida de la función de predicción es del mismo tipo que la columna TARGET de la instrucción CREATE MODEL.

1. Ha entrenado el modelo con datos anteriores al 01/01/2020, por lo que ahora usa la función de predicción en el conjunto de pruebas. La siguiente consulta muestra las predicciones sobre si los clientes que se inscribieron después de 01-01-2020 se darán de baja o no.

```
SELECT
    phone,
    ml_fn_customer_churn_auto(
        state,
        account_length,
        area_code,
        total_charge / account_length,
        cust_serv_calls / account_length
    ) AS active
FROM
    customer_activity
WHERE
    record_date > '2020-01-01';
```

2. El siguiente ejemplo utiliza la misma función de predicción para un caso de uso diferente. En este caso, Amazon Redshift predice la proporción de usuarios que se han dado de baja y los que no lo han hecho entre clientes de diferentes estados en los que la fecha de registro es posterior al 01-01-2020.

```
WITH predicted AS (
    SELECT
        state,
        ml_fn_customer_churn_auto(
            state,
            account_length,
```

```
            area_code,
            total_charge / account_length,
            cust_serv_calls / account_length
        ) :: varchar(6) AS active
    FROM
        customer_activity
    WHERE
        record_date > '2020-01-01'
)
SELECT
    state,
    SUM(
        CASE
            WHEN active = 'True.' THEN 1
            ELSE 0
        END
    ) AS churners,
    SUM(
        CASE
            WHEN active = 'False.' THEN 1
            ELSE 0
        END
    ) AS nonchurners,
    COUNT(*) AS total_per_state
FROM
    predicted
GROUP BY
    state
ORDER BY
    state;
```

3. El siguiente ejemplo utiliza la función de predicción para el caso de uso consistente en predecir el porcentaje de clientes que se dan de baja en un estado. En este caso, Amazon Redshift predice el porcentaje de deserción con una fecha de registro posterior al 01/01/2020.

```
WITH predicted AS (
    SELECT
        state,
        ml_fn_customer_churn_auto(
            state,
            account_length,
            area_code,
            total_charge / account_length,
```

```
            cust_serv_calls / account_length
        ) :: varchar(6) AS active
    FROM
        customer_activity
    WHERE
        record_date > '2020-01-01'
)
SELECT
    state,
    CAST((CAST((SUM(
        CASE
            WHEN active = 'True.' THEN 1
            ELSE 0
        END
    )) AS FLOAT) / CAST(COUNT(*) AS FLOAT)) AS DECIMAL (3, 2)) AS pct_churn,
    COUNT(*) AS total_customers_per_state
FROM
    predicted
GROUP BY
    state
ORDER BY
    3 DESC;
```

## Temas relacionados de

Para obtener más información sobre Amazon Redshift ML, consulte la siguiente documentación:

- Costos de utilizar Amazon Redshift ML
- Comando CREATE MODEL
- Función EXPLAIN_MODEL

Para obtener más información sobre machine learning, consulte la siguiente documentación:

- Información general sobre machine learning
- Machine learning para principiantes y expertos
- ¿Qué es la equidad y la explicabilidad del modelo para las predicciones de machine learning?

# Tutorial: Creación de modelos de agrupación en clústeres k-means

En este tutorial, utilizará Amazon Redshift ML para crear, entrenar e implementar un modelo de machine learning basado en el algoritmo k-means. Este algoritmo resuelve problemas en clústeres en los que desea descubrir agrupaciones en los datos. K-means ayuda a agrupar datos que aún no se han etiquetado. Para obtener más información sobre la agrupación en clústeres k-means, consulte Funcionamiento de la agrupación en clústeres de k-means en la Guía para desarrolladores de Amazon SageMaker.

Utilizará una operación CREATE MODEL para crear un modelo k-means a partir de un clúster de Amazon Redshift. Se puede utilizar un comando CREATE MODEL para exportar datos de entrenamiento, entrenar un modelo, importar ese modelo y preparar una función de predicción de Amazon Redshift. Utilice la operación CREATE MODEL para especificar los datos de entrenamiento en forma de tabla o con la instrucción SELECT.

En este tutorial, utilizará k-means en el conjunto de datos de Global Database of Events, Language, and Tone (GDELT), que monitoriza las noticias en todo el mundo y los datos que se almacenan segundo a segundo cada día. k-means agrupa eventos que tienen tonos, actores o ubicaciones similares. Los datos se almacenan a modo de varios archivos en Amazon Simple Storage Service en dos carpetas diferentes. Las carpetas son históricas (abarcan los años 1979-2013) y hay actualizaciones diarias a partir del año 2013. En este ejemplo, usamos el formato histórico y traemos datos de 1979.

## Ejemplos de casos de uso

Puede resolver otros problemas de agrupamiento en clústeres con Amazon Redshift ML, como el agrupamiento de clientes que tienen hábitos de visualización similares en un servicio de streaming. También puede usar Redshift ML para predecir la cantidad óptima de centros de envío para un servicio de entrega.

Tareas

- Requisitos previos
- Paso 1: Cargar los datos desde Amazon S3 en Amazon Redshift
- Paso 2: Crear el modelo de machine learning
- Paso 3: Realizar predicciones con el modelo

## Requisitos previos

Para completar este tutorial, debe realizar el procedimiento de configuración administrativa de Amazon Redshift ML.

## Paso 1: Cargar los datos desde Amazon S3 en Amazon Redshift

1. Use el editor de consultas de Amazon Redshift v2 para ejecutar la siguiente consulta. La consulta descarta la tabla `gdelt_data` en el esquema público si existe y crea una tabla con el mismo nombre en el esquema público.

```
DROP TABLE IF EXISTS gdelt_data CASCADE;

CREATE TABLE gdelt_data (
    GlobalEventId bigint,
    SqlDate bigint,
    MonthYear bigint,
    Year bigint,
    FractionDate double precision,
    Actor1Code varchar(256),
    Actor1Name varchar(256),
    Actor1CountryCode varchar(256),
    Actor1KnownGroupCode varchar(256),
    Actor1EthnicCode varchar(256),
    Actor1Religion1Code varchar(256),
    Actor1Religion2Code varchar(256),
    Actor1Type1Code varchar(256),
    Actor1Type2Code varchar(256),
    Actor1Type3Code varchar(256),
    Actor2Code varchar(256),
    Actor2Name varchar(256),
    Actor2CountryCode varchar(256),
    Actor2KnownGroupCode varchar(256),
    Actor2EthnicCode varchar(256),
    Actor2Religion1Code varchar(256),
    Actor2Religion2Code varchar(256),
    Actor2Type1Code varchar(256),
    Actor2Type2Code varchar(256),
    Actor2Type3Code varchar(256),
    IsRootEvent bigint,
    EventCode bigint,
    EventBaseCode bigint,
    EventRootCode bigint,
```

```
    QuadClass bigint,
    GoldsteinScale double precision,
    NumMentions bigint,
    NumSources bigint,
    NumArticles bigint,
    AvgTone double precision,
    Actor1Geo_Type bigint,
    Actor1Geo_FullName varchar(256),
    Actor1Geo_CountryCode varchar(256),
    Actor1Geo_ADM1Code varchar(256),
    Actor1Geo_Lat double precision,
    Actor1Geo_Long double precision,
    Actor1Geo_FeatureID bigint,
    Actor2Geo_Type bigint,
    Actor2Geo_FullName varchar(256),
    Actor2Geo_CountryCode varchar(256),
    Actor2Geo_ADM1Code varchar(256),
    Actor2Geo_Lat double precision,
    Actor2Geo_Long double precision,
    Actor2Geo_FeatureID bigint,
    ActionGeo_Type bigint,
    ActionGeo_FullName varchar(256),
    ActionGeo_CountryCode varchar(256),
    ActionGeo_ADM1Code varchar(256),
    ActionGeo_Lat double precision,
    ActionGeo_Long double precision,
    ActionGeo_FeatureID bigint,
    DATEADDED bigint
);
```

2. La siguiente consulta carga los datos de muestra en la tabla `gdelt_data`.

```
COPY gdelt_data
FROM 's3://gdelt-open-data/events/1979.csv'
REGION 'us-east-1'
IAM_ROLE default
CSV
DELIMITER '\t';
```

### Examinar los datos de entrenamiento (opcional)

Para ver en qué datos se entrenará su modelo, utilice la siguiente consulta.

```
SELECT
    AvgTone,
    EventCode,
    NumArticles,
    Actor1Geo_Lat,
    Actor1Geo_Long,
    Actor2Geo_Lat,
    Actor2Geo_Long
FROM
    gdelt_data LIMIT 100;
```

## Paso 2: Crear el modelo de machine learning

El siguiente ejemplo utiliza el comando CREATE MODEL para crear un modelo que agrupa los datos en siete clústeres. El valor K es el número de clústeres en los que se dividen los puntos de datos. El modelo clasifica los puntos de datos en clústeres en los que los puntos de datos son más similares entre sí. A través de la agrupación en clústeres de los puntos de datos en grupos, el algoritmo k-means determina de forma iterativa el mejor centro del clúster. A continuación, el algoritmo asigna cada punto de datos al centro del clúster más cercano. Los miembros más cercanos al mismo centro del clúster pertenecen al mismo grupo. Los miembros de un grupo son lo más parecido posible a otros miembros del mismo grupo y lo más diferente posible de los miembros de otros grupos. El valor K es subjetivo y depende de los métodos que miden las similitudes entre los puntos de datos. Puede cambiar el valor K para suavizar los tamaños de los clústeres si están distribuidos de manera desigual.

En el siguiente ejemplo, reemplace amzn-s3-demo-bucket por el bucket de Amazon S3 propio.

```
CREATE MODEL news_data_clusters
FROM
    (
        SELECT
            AvgTone,
            EventCode,
            NumArticles,
            Actor1Geo_Lat,
            Actor1Geo_Long,
            Actor2Geo_Lat,
            Actor2Geo_Long
        FROM
            gdelt_data
    ) FUNCTION news_monitoring_cluster
```

```
    IAM_ROLE default
    AUTO OFF
    MODEL_TYPE KMEANS
    PREPROCESSORS 'none'
    HYPERPARAMETERS DEFAULT
    EXCEPT
    (K '7')
    SETTINGS (S3_BUCKET 'amzn-s3-demo-bucket');
```

Comprobar el estado del entrenamiento del modelo (opcional)

Puede utilizar el comando SHOW MODEL para saber cuándo está listo el modelo.

Para comprobar el estado del modelo, utilice la siguiente operación SHOW MODEL y determine si `Model State` es `Ready`.

```
SHOW MODEL NEWS_DATA_CLUSTERS;
```

Cuando el modelo esté listo, la salida de la operación anterior debería mostrar que `Model State` es `Ready`. A continuación se muestra un ejemplo de la salida de la operación SHOW MODEL.

```
+--------------------------
+-----------------------------------------------------------------------------------------------
+
|        Model Name        |
 news_data_clusters                                  |
+--------------------------
+-----------------------------------------------------------------------------------------------
+
|       Schema Name        |                                                 public
                                        |
|          Owner           |                                                awsuser
                                        |
|      Creation Time       |                                      Fri, 17.06.2022
 16:32:19                                    |
|       Model State        |                                                 READY
                                        |
|        train:msd         |                                          2973.822754
                                        |
|      train:progress      |                                           100.000000
                                        |
|     train:throughput     |                                         237114.875000
                                        |
```

```
|      Estimated Cost      |                                                    0.004983
                                        |
|                          |
                                        |
|      TRAINING DATA:      |
                                        |
|          Query           | SELECT AVGTONE, EVENTCODE, NUMARTICLES, ACTOR1GEO_LAT,
 ACTOR1GEO_LONG, ACTOR2GEO_LAT, ACTOR2GEO_LONG |
|                          |                                                FROM GDELT_DATA
                                        |
|                          |
                                        |
|       PARAMETERS:        |
                                        |
|        Model Type        |                                                     kmeans
                                        |
|    Training Job Name     |
 redshiftml-20220617163219978978-kmeans                                |
|      Function Name       |
 news_monitoring_cluster                                      |
|   Function Parameters    |       avgtone eventcode numarticles actor1geo_lat
 actor1geo_long actor2geo_lat actor2geo_long        |
| Function Parameter Types |                             float8 int8 int8 float8 float8
 float8 float8                          |
|         IAM Role         |                                        default-aws-iam-
role                                      |
|        S3 Bucket         |                                           amzn-s3-demo-
bucket                                    |
|       Max Runtime        |                                                5400
                                        |
|                          |
                                        |
|     HYPERPARAMETERS:     |
                                        |
|       feature_dim        |                                                 7
                                        |
|            k             |                                                 7
                                        |
+--------------------------
+-----------------------------------------------------------------------------------------------
+
```

## Paso 3: Realizar predicciones con el modelo

### Identificar los clústeres

Puede encontrar agrupaciones discretas identificadas en los datos por su modelo, también conocidas como clústeres. Un clúster es el conjunto de puntos de datos que está más cerca de su centro de clúster que de cualquier otro centro de clúster. Dado que el valor K representa el número de clústeres del modelo, también representa el número de centros de clúster. La siguiente consulta identifica los clústeres mostrando el clúster asociado a cada `globaleventid`.

```
SELECT
    globaleventid,
    news_monitoring_cluster (
        AvgTone,
        EventCode,
        NumArticles,
        Actor1Geo_Lat,
        Actor1Geo_Long,
        Actor2Geo_Lat,
        Actor2Geo_Long
    ) AS cluster
FROM
    gdelt_data;
```

### Comprobar la distribución de los datos

Puede comprobar la distribución de los datos entre los clústeres para ver si el valor K que eligió ha provocado que los datos se distribuyeran de manera uniforme. Utilice la siguiente consulta para determinar si los datos están distribuidos de manera uniforme en los clústeres.

```
SELECT
    events_cluster,
    COUNT(*) AS nbr_events
FROM
    (
        SELECT
            globaleventid,
            news_monitoring_cluster(
                AvgTone,
                EventCode,
                NumArticles,
                Actor1Geo_Lat,
```

```
                Actor1Geo_Long,
                Actor2Geo_Lat,
                Actor2Geo_Long
            ) AS events_cluster
        FROM
            gdelt_data
    )
GROUP BY
    1;
```

Puede cambiar el valor K para suavizar los tamaños de los clústeres si no están distribuidos de manera uniforme.

Determinar los centros de clúster

Un punto de datos está más cerca de su centro de clúster que de cualquier otro centro de clúster. Por lo tanto, encontrar los centros de clúster le ayuda a definir los clústeres.

Ejecute la siguiente consulta para determinar los centros de clúster en función del número de artículos por código de evento.

```
SELECT
    news_monitoring_cluster (
        AvgTone,
        EventCode,
        NumArticles,
        Actor1Geo_Lat,
        Actor1Geo_Long,
        Actor2Geo_Lat,
        Actor2Geo_Long
    ) AS events_cluster,
    eventcode,
    SUM(numArticles) AS numArticles
FROM
    gdelt_data
GROUP BY
    1,
    2;
```

Mostrar información sobre puntos de datos en un clúster

Utilice la siguiente consulta para devolver los datos de los puntos asignados al quinto clúster. Los artículos seleccionados deben tener dos actores.

```
SELECT
    news_monitoring_cluster (
        AvgTone,
        EventCode,
        NumArticles,
        Actor1Geo_Lat,
        Actor1Geo_Long,
        Actor2Geo_Lat,
        Actor2Geo_Long
    ) AS events_cluster,
    eventcode,
    actor1name,
    actor2name,
    SUM(numarticles) AS totalarticles
FROM
    gdelt_data
WHERE
    events_cluster = 5
    AND actor1name <> ' '
    AND actor2name <> ' '
GROUP BY
    1,
    2,
    3,
    4
ORDER BY
    5 desc;
```

Mostrar datos sobre eventos con actores del mismo código étnico

La siguiente consulta cuenta la cantidad de artículos escritos sobre eventos de tono positivo. La consulta también requiere que los dos actores tengan el mismo código étnico y devuelve a qué grupo está asignado cada evento.

```
SELECT
    news_monitoring_cluster (
        AvgTone,
        EventCode,
        NumArticles,
        Actor1Geo_Lat,
        Actor1Geo_Long,
        Actor2Geo_Lat,
        Actor2Geo_Long
```

```
    ) AS events_cluster,
    SUM(numarticles) AS total_articles,
    eventcode AS event_code,
    Actor1EthnicCode AS ethnic_code
FROM
    gdelt_data
WHERE
    Actor1EthnicCode = Actor2EthnicCode
    AND Actor1EthnicCode <> ' '
    AND Actor2EthnicCode <> ' '
    AND AvgTone > 0
GROUP BY
    1,
    3,
    4
HAVING
    (total_articles) > 4
ORDER BY
    1,
    2 ASC;
```

### Temas relacionados de

Para obtener más información sobre Amazon Redshift ML, consulte la siguiente documentación:

- Costes de utilizar Amazon Redshift ML
- Operación CREATE MODEL
- Función EXPLAIN_MODEL

Para obtener más información sobre machine learning, consulte la siguiente documentación:

- Información general sobre machine learning
- Machine learning para principiantes y expertos
- ¿Qué es la equidad y la explicabilidad del modelo para las predicciones de machine learning?

## Tutorial: Creación de modelos de clasificación multiclase

En este tutorial, utilizará Amazon Redshift ML para crear un modelo de machine learning que resuelva los problemas de clasificación multiclase. El algoritmo de clasificación multiclase clasifica

los puntos de datos en una de tres o más clases. A continuación, implementará consultas mediante la función SQL que genera el comando CREATE MODEL.

Se puede utilizar un comando CREATE MODEL para exportar datos de entrenamiento, entrenar un modelo, importar ese modelo y preparar una función de predicción de Amazon Redshift. Utilice la operación CREATE MODEL para especificar los datos de entrenamiento en forma de tabla o con la instrucción SELECT.

Para seguir el tutorial, utilice el conjunto de datos público E-Commerce Sales Forecast (Previsión de ventas de comercio electrónico), que incluye datos de ventas de un minorista en línea del Reino Unido. El modelo que genere se dirigirá a los clientes más activos para un programa especial de fidelización de clientes. Con la clasificación multiclase, puede usar el modelo para predecir cuántos meses estará activo un cliente durante un período de 13 meses. La función de predicción designa a los clientes que se prevé que estarán activos durante 7 o más meses para la admisión al programa.

## Ejemplos de casos de uso

Puede resolver otros problemas de clasificación multiclase con Amazon Redshift ML, como predecir el producto más vendido de una línea de productos. También puede predecir qué fruta contiene una imagen, por ejemplo, seleccionar manzanas, peras o naranjas.

Tareas

- Requisitos previos
- Paso 1: Cargar los datos desde Amazon S3 en Amazon Redshift
- Paso 2: Crear el modelo de machine learning
- Paso 3: Realizar predicciones con el modelo

## Requisitos previos

Para completar este tutorial, debe realizar el procedimiento de configuración administrativa de Amazon Redshift ML.

## Paso 1: Cargar los datos desde Amazon S3 en Amazon Redshift

Use el editor de consultas de Amazon Redshift v2 para ejecutar las siguientes consultas. Estas consultas cargan los datos de muestra en Amazon Redshift.

1. En la siguiente consulta, se crea una tabla llamada `ecommerce_sales`.

```
CREATE TABLE IF NOT EXISTS ecommerce_sales (
    invoiceno VARCHAR(30),
    stockcode VARCHAR(30),
    description VARCHAR(60),
    quantity DOUBLE PRECISION,
    invoicedate VARCHAR(30),
    unitprice DOUBLE PRECISION,
    customerid BIGINT,
    country VARCHAR(25)
);
```

2. La siguiente consulta copia los datos de muestra del conjunto de datos E-Commerce Sales Forecast (Previsión de ventas de comercio electrónico) en la tabla `ecommerce_sales`.

```
COPY ecommerce_sales
FROM
    's3://redshift-ml-multiclass/ecommerce_data.txt'
IAM_ROLE default
DELIMITER '\t'
IGNOREHEADER 1
REGION 'us-east-1'
MAXERROR 100;
```

### Dividir los datos

Al crear un modelo en Amazon Redshift ML, SageMaker AI divide automáticamente los datos en conjuntos de entrenamiento y de prueba para que SageMaker AI pueda determinar la precisión del modelo. Al dividir manualmente los datos en este paso, podrá verificar la precisión del modelo asignando un conjunto de predicciones adicional.

Utilice la siguiente instrucción SQL para dividir los datos en tres conjuntos para el entrenamiento, la validación y la predicción.

```
--creates table with all data
CREATE TABLE ecommerce_sales_data AS (
    SELECT
        t1.stockcode,
        t1.description,
        t1.invoicedate,
        t1.customerid,
```

```
        t1.country,
        t1.sales_amt,
        CAST(RANDOM() * 100 AS INT) AS data_group_id
    FROM
        (
            SELECT
                stockcode,
                description,
                invoicedate,
                customerid,
                country,
                SUM(quantity * unitprice) AS sales_amt
            FROM
                ecommerce_sales
            GROUP BY
                1,
                2,
                3,
                4,
                5
        ) t1
);

--creates training set
CREATE TABLE ecommerce_sales_training AS (
    SELECT
        a.customerid,
        a.country,
        a.stockcode,
        a.description,
        a.invoicedate,
        a.sales_amt,
        (b.nbr_months_active) AS nbr_months_active
    FROM
        ecommerce_sales_data a
        INNER JOIN (
            SELECT
                customerid,
                COUNT(
                    DISTINCT(
                        DATE_PART(y, CAST(invoicedate AS DATE)) || '-' || LPAD(
                            DATE_PART(mon, CAST(invoicedate AS DATE)),
                            2,
                            '00'
```

```
                        )
                    )
                ) AS nbr_months_active
            FROM
                ecommerce_sales_data
            GROUP BY
                1
        ) b ON a.customerid = b.customerid
    WHERE
        a.data_group_id < 80
);

--creates validation set
CREATE TABLE ecommerce_sales_validation AS (
    SELECT
        a.customerid,
        a.country,
        a.stockcode,
        a.description,
        a.invoicedate,
        a.sales_amt,
        (b.nbr_months_active) AS nbr_months_active
    FROM
        ecommerce_sales_data a
        INNER JOIN (
            SELECT
                customerid,
                COUNT(
                    DISTINCT(
                        DATE_PART(y, CAST(invoicedate AS DATE)) || '-' || LPAD(
                            DATE_PART(mon, CAST(invoicedate AS DATE)),
                            2,
                            '00'
                        )
                    )
                ) AS nbr_months_active
            FROM
                ecommerce_sales_data
            GROUP BY
                1
        ) b ON a.customerid = b.customerid
    WHERE
        a.data_group_id BETWEEN 80
        AND 90
```

```
);

--creates prediction set
CREATE TABLE ecommerce_sales_prediction AS (
    SELECT
        customerid,
        country,
        stockcode,
        description,
        invoicedate,
        sales_amt
    FROM
        ecommerce_sales_data
    WHERE
        data_group_id > 90);
```

## Paso 2: Crear el modelo de machine learning

En este paso, utilizará la instrucción CREATE MODEL para crear el modelo de machine learning mediante la clasificación multiclase.

La siguiente consulta crea el modelo de clasificación multiclase con el conjunto de entrenamiento mediante la operación CREATE MODEL. Reemplace amzn-s3-demo-bucket por el bucket de Amazon S3 propio.

```
CREATE MODEL ecommerce_customer_activity
FROM
    (
        SELECT
            customerid,
            country,
            stockcode,
            description,
            invoicedate,
            sales_amt,
            nbr_months_active
        FROM
            ecommerce_sales_training
    ) TARGET nbr_months_active FUNCTION predict_customer_activity IAM_ROLE default
 PROBLEM_TYPE MULTICLASS_CLASSIFICATION SETTINGS (
        S3_BUCKET 'amzn-s3-demo-bucket',
        S3_GARBAGE_COLLECT OFF
```

```
    );
```

En esta consulta, se especifica el tipo de problema como `Multiclass_Classification`. El objetivo que predice para el modelo es `nbr_months_active`. Cuando SageMaker AI termina de entrenar el modelo, crea la función `predict_customer_activity`, que utilizará para hacer predicciones en Amazon Redshift.

Comprobar el estado del entrenamiento del modelo (opcional)

Puede utilizar el comando SHOW MODEL para saber cuándo está listo el modelo.

Utilice la siguiente consulta para devolver varias métricas del modelo, incluidos el estado y la precisión del modelo.

```
SHOW MODEL ecommerce_customer_activity;
```

Cuando el modelo esté listo, la salida de la operación anterior debería mostrar que `Model State` es `Ready`. A continuación se muestra un ejemplo de la salida de la operación SHOW MODEL.

```
+--------------------------
+-----------------------------------------------------------------------------------------------
+
|        Model Name        |
 ecommerce_customer_activity                                |
+--------------------------
+-----------------------------------------------------------------------------------------------
+
|       Schema Name        |                                          public
                                  |
|          Owner           |                                          awsuser
                                  |
|      Creation Time       |                                Fri, 17.06.2022 19:02:15
                                  |
|       Model State        |                                           READY
                                  |
|   Training Job Status    |
 MaxAutoMLJobRuntimeReached                                 |
|   validation:accuracy    |                                        0.991280
                                  |
|      Estimated Cost      |                                        7.897689
                                  |
```

```
|                          |                                                
```

```
SELECT
    customerid,
    predict_customer_activity(
        customerid,
        country,
        stockcode,
        description,
        invoicedate,
        sales_amt
    ) AS predicted_months_active
FROM
    ecommerce_sales_prediction
WHERE
    predicted_months_active >= 7
GROUP BY
    1,
    2
LIMIT
    10;
```

Ejecutar consultas de predicción contra los datos de validación (opcional)

Ejecute las siguientes consultas de predicción con los datos de validación para comprobar el nivel de precisión del modelo.

```
SELECT
    CAST(SUM(t1.match) AS decimal(7, 2)) AS predicted_matches,
    CAST(SUM(t1.nonmatch) AS decimal(7, 2)) AS predicted_non_matches,
    CAST(SUM(t1.match + t1.nonmatch) AS decimal(7, 2)) AS total_predictions,
    predicted_matches / total_predictions AS pct_accuracy
FROM
    (
        SELECT
            customerid,
            country,
            stockcode,
            description,
            invoicedate,
            sales_amt,
            nbr_months_active,
            predict_customer_activity(
                customerid,
                country,
```

```
                stockcode,
                description,
                invoicedate,
                sales_amt
            ) AS predicted_months_active,
            CASE
                WHEN nbr_months_active = predicted_months_active THEN 1
                ELSE 0
            END AS match,
            CASE
                WHEN nbr_months_active <> predicted_months_active THEN 1
                ELSE 0
            END AS nonmatch
        FROM
            ecommerce_sales_validation
    )t1;
```

Prediga cuántos clientes no ingresan (opcional)

La siguiente consulta compara la cantidad de clientes que se prevé que estarán activos solo durante 5 o 6 meses. El modelo predice que estos clientes se quedarán fuera del programa de fidelización. A continuación, la consulta compara la cantidad que se quedan fuera del programa con el número que se prevé que cumpla los requisitos para el programa de fidelización. Esta consulta podría usarse para tomar una decisión sobre si reducir el umbral del programa de fidelización. También puede determinar si hay una cantidad significativa de clientes que se prevé que se quedarán fuera del programa por poco. A continuación, puede alentar a esos clientes a aumentar su actividad para poder entrar en el programa de fidelización.

```
SELECT
    predict_customer_activity(
        customerid,
        country,
        stockcode,
        description,
        invoicedate,
        sales_amt
    ) AS predicted_months_active,
    COUNT(customerid)
FROM
    ecommerce_sales_prediction
WHERE
    predicted_months_active BETWEEN 5 AND 6
```

```
GROUP BY
    1
ORDER BY
    1 ASC
LIMIT
    10)
UNION
(SELECT
      NULL AS predicted_months_active,
    COUNT (customerid)
FROM
    ecommerce_sales_prediction
WHERE
    predict_customer_activity(
        customerid,
        country,
        stockcode,
        description,
        invoicedate,
        sales_amt
    ) >=7);
```

### Temas relacionados de

Para obtener más información sobre Amazon Redshift ML, consulte la siguiente documentación:

- Costos de utilizar Amazon Redshift ML
- Operación CREATE MODEL
- Función EXPLAIN_MODEL

Para obtener más información sobre machine learning, consulte la siguiente documentación:

- Información general sobre machine learning
- Machine learning para principiantes y expertos
- ¿Qué es la equidad y la explicabilidad del modelo para las predicciones de machine learning?

## Tutorial: Creación de modelos XGBoost

En este tutorial, creará un modelo con datos de Amazon S3 y ejecutará consultas de predicción con el modelo mediante Amazon Redshift ML. El algoritmo XGBoost es una implementación

optimizada del algoritmo de árboles aumentados. XGBoost maneja más tipos de datos, relaciones y distribuciones que otros algoritmos de árboles aumentados. Puede usar XGBoost para problemas de regresión, de clasificación binaria, clasificación multiclase) y de ranking. Para obtener más información sobre el algoritmo XGBoost, consulte Algoritmo XGBoost en la Guía para desarrolladores de Amazon SageMaker AI.

Actualmente,la operación `CREATE MODEL` de Amazon Redshift ML con la opción `AUTO OFF` es compatible con XGBoost como `MODEL_TYPE`. Puede proporcionar información relevante, como el objetivo y los hiperparámetros, como parte de `CREATE MODEL`, en función del caso de uso.

En este tutorial, se utiliza el conjunto de datos de autenticación de billetes, que es un problema de clasificación binaria para predecir si un billete determinado es auténtico o falso.

## Ejemplos de casos de uso

Puede resolver otros problemas de clasificación binaria con Amazon Redshift ML, como predecir si un paciente está sano o tiene una enfermedad. También puede predecir si un correo electrónico es spam o no.

Tareas

- Requisitos previos
- Paso 1: Cargar los datos desde Amazon S3 en Amazon Redshift
- Paso 2: Crear el modelo de machine learning
- Paso 3: Realizar predicciones con el modelo

## Requisitos previos

Para completar este tutorial, debe realizar el procedimiento de configuración administrativa de Amazon Redshift ML.

## Paso 1: Cargar los datos desde Amazon S3 en Amazon Redshift

Use el editor de consultas de Amazon Redshift v2 para ejecutar las siguientes consultas.

La siguiente consulta crea dos tablas, carga los datos de Amazon S3 y divide los datos en un conjunto de entrenamiento y un conjunto de pruebas. Utilizará el conjunto de entrenamiento para entrenar su modelo y crear la función de predicción. A continuación, probará la función de predicción en el conjunto de pruebas.

```
--create training set table
CREATE TABLE banknoteauthentication_train(
    variance FLOAT,
    skewness FLOAT,
    curtosis FLOAT,
    entropy FLOAT,
    class INT
);

--Load into training table
COPY banknoteauthentication_train
FROM
    's3://redshiftbucket-ml-sagemaker/banknote_authentication/train_data/' IAM_ROLE
 default REGION 'us-west-2' IGNOREHEADER 1 CSV;

--create testing set table
CREATE TABLE banknoteauthentication_test(
    variance FLOAT,
    skewness FLOAT,
    curtosis FLOAT,
    entropy FLOAT,
    class INT
);

--Load data into testing table
COPY banknoteauthentication_test
FROM
    's3://redshiftbucket-ml-sagemaker/banknote_authentication/test_data/'
    IAM_ROLE default
    REGION 'us-west-2'
    IGNOREHEADER 1
    CSV;
```

## Paso 2: Crear el modelo de machine learning

La siguiente consulta crea el modelo XGBoost en Amazon Redshift ML a partir del conjunto de entrenamiento que creó en el paso anterior. Reemplace `amzn-s3-demo-bucket` con su propio `S3_BUCKET`, que almacenará sus conjuntos de datos de entrada y otros artefactos de Redshift ML.

```
CREATE MODEL model_banknoteauthentication_xgboost_binary
FROM
    banknoteauthentication_train
```

```
    TARGET class
    FUNCTION func_model_banknoteauthentication_xgboost_binary
    IAM_ROLE default
    AUTO OFF
    MODEL_TYPE xgboost
    OBJECTIVE 'binary:logistic'
    PREPROCESSORS 'none'
    HYPERPARAMETERS DEFAULT
EXCEPT(NUM_ROUND '100')
SETTINGS(S3_BUCKET 'amzn-s3-demo-bucket');
```

Comprobar el estado del entrenamiento del modelo (opcional)

Puede utilizar el comando SHOW MODEL para saber cuándo está listo el modelo.

Utilice la siguiente consulta para supervisar el estado del entrenamiento del modelo.

```
SHOW MODEL model_banknoteauthentication_xgboost_binary;
```

Si el modelo es READY, la operación SHOW MODEL también proporciona la métrica `train:error`, como se muestra en el siguiente ejemplo de la salida. La métrica `train:error` es una medida de precisión de su modelo que mide hasta seis decimales. Un valor de 0 es más preciso y un valor de 1 es menos preciso.

```
+--------------------------+-------------------------------------------------+
|        Model Name        |   model_banknoteauthentication_xgboost_binary   |
+--------------------------+-------------------------------------------------+
| Schema Name              | public                                          |
| Owner                    | awsuser                                         |
| Creation Time            | Tue, 21.06.2022 19:07:35                        |
| Model State              | READY                                           |
| train:error              |                                        0.000000 |
| Estimated Cost           |                                        0.006197 |
|                          |                                                 |
| TRAINING DATA:           |                                                 |
| Query                    | SELECT *                                        |
|                          | FROM "BANKNOTEAUTHENTICATION_TRAIN"             |
| Target Column            | CLASS                                           |
|                          |                                                 |
| PARAMETERS:              |                                                 |
| Model Type               | xgboost                                         |
| Training Job Name        | redshiftml-20220621190735686935-xgboost         |
| Function Name            | func_model_banknoteauthentication_xgboost_binary |
```

```
| Function Parameters      | variance skewness curtosis entropy                |
| Function Parameter Types | float8 float8 float8 float8                       |
| IAM Role                 | default-aws-iam-role                              |
| S3 Bucket                | amzn-s3-demo-bucket                               |
| Max Runtime              |                                              5400 |
|                          |                                                   |
| HYPERPARAMETERS:         |                                                   |
| num_round                |                                               100 |
| objective                | binary:logistic                                   |
+--------------------------+---------------------------------------------------+
```

## Paso 3: Realizar predicciones con el modelo

### Comprobar la precisión del modelo

La siguiente consulta de predicción utiliza la función de predicción creada en el paso anterior para comprobar la precisión del modelo. Ejecute esta consulta en el conjunto de pruebas para asegurarse de que el modelo no se corresponde demasiado con el conjunto de entrenamiento. Esta estrecha correspondencia también se conoce como sobreajuste, y el sobreajuste podría hacer que el modelo hiciera predicciones poco fiables.

```
WITH predict_data AS (
    SELECT
        class AS label,
        func_model_banknoteauthentication_xgboost_binary (variance, skewness, curtosis,
 entropy) AS predicted,
        CASE
            WHEN label IS NULL THEN 0
            ELSE label
        END AS actual,
        CASE
            WHEN actual = predicted THEN 1 :: INT
            ELSE 0 :: INT
        END AS correct
    FROM
        banknoteauthentication_test
),
aggr_data AS (
    SELECT
        SUM(correct) AS num_correct,
        COUNT(*) AS total
```

```
    FROM
        predict_data
)
SELECT
    (num_correct :: FLOAT / total :: FLOAT) AS accuracy
FROM
    aggr_data;
```

Predecir la cantidad de billetes auténticos y falsos

La siguiente consulta de predicción devuelve la cantidad prevista de billetes auténticos y falsos en el conjunto de pruebas.

```
WITH predict_data AS (
    SELECT
        func_model_banknoteauthentication_xgboost_binary(variance, skewness, curtosis,
 entropy) AS predicted
    FROM
        banknoteauthentication_test
)
SELECT
    CASE
        WHEN predicted = '0' THEN 'Original banknote'
        WHEN predicted = '1' THEN 'Counterfeit banknote'
        ELSE 'NA'
    END AS banknote_authentication,
    COUNT(1) AS count
FROM
    predict_data
GROUP BY
    1;
```

Encontrar la observación promedio de un billete auténtico y uno falso

La siguiente consulta de predicción devuelve el valor promedio de cada característica para los billetes que se predice que son auténticos y falsos en el conjunto de prueba.

```
WITH predict_data AS (
    SELECT
        func_model_banknoteauthentication_xgboost_binary(variance, skewness, curtosis,
 entropy) AS predicted,
          variance,
          skewness,
```

```
          curtosis,
          entropy
    FROM
        banknoteauthentication_test
)
SELECT
    CASE
        WHEN predicted = '0' THEN 'Original banknote'
        WHEN predicted = '1' THEN 'Counterfeit banknote'
        ELSE 'NA'
    END AS banknote_authentication,
    TRUNC(AVG(variance), 2) AS avg_variance,
    TRUNC(AVG(skewness), 2) AS avg_skewness,
    TRUNC(AVG(curtosis), 2) AS avg_curtosis,
    TRUNC(AVG(entropy), 2) AS avg_entropy
FROM
    predict_data
GROUP BY
    1
ORDER BY
    2;
```

### Temas relacionados de

Para obtener más información sobre Amazon Redshift ML, consulte la siguiente documentación:

- Costos de utilizar Amazon Redshift ML
- Operación CREATE MODEL
- Función EXPLAIN_MODEL

Para obtener más información sobre machine learning, consulte la siguiente documentación:

- Información general sobre machine learning
- Machine learning para principiantes y expertos
- ¿Qué es la equidad y la explicabilidad del modelo para las predicciones de machine learning?

## Tutorial: Creación de modelos de regresión

En este tutorial, utilizará Amazon Redshift ML para crear un modelo de regresión de machine learning y ejecutar consultas de predicción en el modelo. Los modelos de regresión le permiten

predecir resultados numéricos, como el precio de una casa o cuántas personas utilizarán el servicio de alquiler de bicicletas de una ciudad. El comando CREATE MODEL se usa en Amazon Redshift con los datos de entrenamiento. A continuación, Amazon Redshift ML compila el modelo, importa el modelo entrenado a Redshift y prepara una función de predicción de SQL. Puede usar la función de predicción en consultas SQL en Amazon Redshift.

En este tutorial, utilizará Amazon Redshift ML para crear un modelo de regresión que prediga el número de personas que utilizan el servicio de bicicletas compartidas de la ciudad de Toronto a cualquier hora del día. Las entradas para el modelo incluyen días festivos y condiciones meteorológicas. Utilizará un modelo de regresión, porque quiere un resultado numérico para este problema.

Puede utilizar el comando CREATE MODEL para exportar datos de entrenamiento, entrenar un modelo y hacer que el modelo esté disponible en Amazon Redshift como una función SQL. Utilice la operación CREATE MODEL para especificar los datos de entrenamiento en forma de tabla o con la instrucción SELECT.

## Ejemplos de casos de uso

Puede resolver otros problemas de regresión con Amazon Redshift ML, como predecir el valor de por vida de un cliente. También puede usar Redshift ML para predecir el precio más rentable y los ingresos resultantes de un producto.

Tareas

- Requisitos previos
- Paso 1: Cargar los datos desde Amazon S3 en Amazon Redshift
- Paso 2: Crear el modelo de machine learning
- Paso 3: Validar el modelo

## Requisitos previos

Para completar este tutorial, debe realizar el procedimiento de configuración administrativa de Amazon Redshift ML.

## Paso 1: Cargar los datos desde Amazon S3 en Amazon Redshift

Use el editor de consultas de Amazon Redshift v2 para ejecutar las siguientes consultas.

1. Debe crear tres tablas para cargar los tres conjuntos de datos públicos en Amazon Redshift. Los conjuntos de datos son los datos del número de usuarios de bicicletas de Toronto, losdatos meteorológicos históricos y los datos históricos de días festivos. Ejecute la siguiente consulta en el editor de consultas de Amazon Redshift para crear las tablas denominadas `ridership`, `weather` y `holiday`.

```
CREATE TABLE IF NOT EXISTS ridership (
    trip_id INT,
    trip_duration_seconds INT,
    trip_start_time timestamp,
    trip_stop_time timestamp,
    from_station_name VARCHAR(50),
    to_station_name VARCHAR(50),
    from_station_id SMALLINT,
    to_station_id SMALLINT,
    user_type VARCHAR(20)
);

CREATE TABLE IF NOT EXISTS weather (
    longitude_x DECIMAL(5, 2),
    latitude_y DECIMAL(5, 2),
    station_name VARCHAR(20),
    climate_id BIGINT,
    datetime_utc TIMESTAMP,
    weather_year SMALLINT,
    weather_month SMALLINT,
    weather_day SMALLINT,
    time_utc VARCHAR(5),
    temp_c DECIMAL(5, 2),
    temp_flag VARCHAR(1),
    dew_point_temp_c DECIMAL(5, 2),
    dew_point_temp_flag VARCHAR(1),
    rel_hum SMALLINT,
    rel_hum_flag VARCHAR(1),
    precip_amount_mm DECIMAL(5, 2),
    precip_amount_flag VARCHAR(1),
    wind_dir_10s_deg VARCHAR(10),
    wind_dir_flag VARCHAR(1),
    wind_spd_kmh VARCHAR(10),
    wind_spd_flag VARCHAR(1),
    visibility_km VARCHAR(10),
    visibility_flag VARCHAR(1),
    stn_press_kpa DECIMAL(5, 2),
```

```
    stn_press_flag VARCHAR(1),
    hmdx SMALLINT,
    hmdx_flag VARCHAR(1),
    wind_chill VARCHAR(10),
    wind_chill_flag VARCHAR(1),
    weather VARCHAR(10)
);

CREATE TABLE IF NOT EXISTS holiday (holiday_date DATE, description VARCHAR(100));
```

2. La siguiente consulta carga los datos de muestra en las tablas que ha creado en el paso anterior.

```
COPY ridership
FROM
    's3://redshift-ml-bikesharing-data/bike-sharing-data/ridership/'
    IAM_ROLE default
    FORMAT CSV
    IGNOREHEADER 1
    DATEFORMAT 'auto'
    TIMEFORMAT 'auto'
    REGION 'us-west-2'
    gzip;

COPY weather
FROM
    's3://redshift-ml-bikesharing-data/bike-sharing-data/weather/'
    IAM_ROLE default
    FORMAT csv
    IGNOREHEADER 1
    DATEFORMAT 'auto'
    TIMEFORMAT 'auto'
    REGION 'us-west-2'
    gzip;

COPY holiday
FROM
    's3://redshift-ml-bikesharing-data/bike-sharing-data/holiday/'
    IAM_ROLE default
    FORMAT csv
    IGNOREHEADER 1
    DATEFORMAT 'auto'
    TIMEFORMAT 'auto'
    REGION 'us-west-2'
```

```
    gzip;
```

3. La siguiente consulta realiza transformaciones en los conjuntos da datos `ridership` y `weather` para eliminar sesgos o anomalías. La eliminación de sesgos y anomalías mejora la precisión del modelo. La consulta simplifica las tablas mediante la creación de dos nuevas vistas denominadas `ridership_view` y `weather_view`.

```
CREATE
OR REPLACE VIEW ridership_view AS
SELECT
    trip_time,
    trip_count,
    TO_CHAR(trip_time, 'hh24') :: INT trip_hour,
    TO_CHAR(trip_time, 'dd') :: INT trip_day,
    TO_CHAR(trip_time, 'mm') :: INT trip_month,
    TO_CHAR(trip_time, 'yy') :: INT trip_year,
    TO_CHAR(trip_time, 'q') :: INT trip_quarter,
    TO_CHAR(trip_time, 'w') :: INT trip_month_week,
    TO_CHAR(trip_time, 'd') :: INT trip_week_day
FROM
    (
        SELECT
            CASE
                WHEN TRUNC(r.trip_start_time) < '2017-07-01' :: DATE THEN
 CONVERT_TIMEZONE(
                    'US/Eastern',
                    DATE_TRUNC('hour', r.trip_start_time)
                )
                ELSE DATE_TRUNC('hour', r.trip_start_time)
            END trip_time,
            COUNT(1) trip_count
        FROM
            ridership r
        WHERE
            r.trip_duration_seconds BETWEEN 60
            AND 60 * 60 * 24
        GROUP BY
            1
    );

CREATE
OR REPLACE VIEW weather_view AS
SELECT
```

```
    CONVERT_TIMEZONE(
        'US/Eastern',
        DATE_TRUNC('hour', datetime_utc)
    ) daytime,
    ROUND(AVG(temp_c)) temp_c,
    ROUND(AVG(precip_amount_mm)) precip_amount_mm
FROM
    weather
GROUP BY
    1;
```

4. La siguiente consulta crea una tabla que combina todos los atributos de entrada relevantes de `ridership_view` y `weather_view` en la tabla `trip_data`.

```
CREATE TABLE trip_data AS
SELECT
    r.trip_time,
    r.trip_count,
    r.trip_hour,
    r.trip_day,
    r.trip_month,
    r.trip_year,
    r.trip_quarter,
    r.trip_month_week,
    r.trip_week_day,
    w.temp_c,
    w.precip_amount_mm,CASE
        WHEN h.holiday_date IS NOT NULL THEN 1
        WHEN TO_CHAR(r.trip_time, 'D') :: INT IN (1, 7) THEN 1
        ELSE 0
    END is_holiday,
    ROW_NUMBER() OVER (
        ORDER BY
            RANDOM()
    ) serial_number
FROM
    ridership_view r
    JOIN weather_view w ON (r.trip_time = w.daytime)
    LEFT OUTER JOIN holiday h ON (TRUNC(r.trip_time) = h.holiday_date);
```

Ver los datos de muestra (opcional)

La siguiente consulta muestra las entradas de la tabla. Puede ejecutar esta operación para asegurarse de que la tabla se creó correctamente.

```
SELECT *
FROM trip_data
LIMIT 5;
```

A continuación se muestra un ejemplo de la salida de la operación anterior.

```
+---------------------+------------+----------+----------+------------+-----------
+--------------+-----------------+---------------+--------+------------------
+------------+---------------+
|      trip_time      | trip_count | trip_hour | trip_day | trip_month | trip_year
 | trip_quarter | trip_month_week | trip_week_day | temp_c | precip_amount_mm |
 is_holiday | serial_number |
+---------------------+------------+----------+----------+------------+-----------
+--------------+-----------------+---------------+--------+------------------
+------------+---------------+
| 2017-03-21 22:00:00 |         47 |        22 |       21 |          3 |        17 |
         1 |               3 |             3 |      1 |                0 |          0 |
             1 |
| 2018-05-04 01:00:00 |         19 |         1 |        4 |          5 |        18 |
         2 |               1 |             6 |     12 |                0 |          0 |
             3 |
| 2018-01-11 10:00:00 |         93 |        10 |       11 |          1 |        18 |
         1 |               2 |             5 |      9 |                0 |          0 |
             5 |
| 2017-10-28 04:00:00 |         20 |         4 |       28 |         10 |        17 |
         4 |               4 |             7 |     11 |                0 |          1 |
             7 |
| 2017-12-31 21:00:00 |         11 |        21 |       31 |         12 |        17 |
         4 |               5 |             1 |    -15 |                0 |          1 |
             9 |
+---------------------+------------+----------+----------+------------+-----------
+--------------+-----------------+---------------+--------+------------------
+------------+---------------+
```

Mostrar la correlación entre los atributos (opcional)

Determinar la correlación le ayuda a medir la fuerza de la asociación entre los atributos. El nivel de asociación puede ayudarle a determinar qué afecta a la salida objetivo. En este tutorial, la salida objetivo es `trip_count`.

La siguiente consulta crea o reemplaza el procedimiento `sp_correlation`. Se utiliza el procedimiento almacenado denominado `sp_correlation` para mostrar la correlación entre un atributo y otros atributos en una tabla de Amazon Redshift.

```
CREATE OR REPLACE PROCEDURE sp_correlation(source_schema_name in varchar(255),
 source_table_name in varchar(255), target_column_name in varchar(255),
 output_temp_table_name inout varchar(255)) AS $$
DECLARE
  v_sql varchar(max);
  v_generated_sql varchar(max);
  v_source_schema_name varchar(255)=lower(source_schema_name);
  v_source_table_name varchar(255)=lower(source_table_name);
  v_target_column_name varchar(255)=lower(target_column_name);
BEGIN
  EXECUTE 'DROP TABLE IF EXISTS ' || output_temp_table_name;
  v_sql = '
SELECT
  ''CREATE temp table '|| output_temp_table_name||' AS SELECT ''|| outer_calculation||
  '' FROM (SELECT COUNT(1) number_of_items, SUM('||v_target_column_name||')
 sum_target, SUM(POW('||v_target_column_name||',2)) sum_square_target, POW(SUM('||
v_target_column_name||'),2) square_sum_target,''||
  inner_calculation||
  '' FROM (SELECT ''||
  column_name||
  '' FROM '||v_source_table_name||'))''
FROM
  (
  SELECT
    DISTINCT
    LISTAGG(outer_calculation,'','') OVER () outer_calculation
    ,LISTAGG(inner_calculation,'','') OVER () inner_calculation
    ,LISTAGG(column_name,'','') OVER () column_name
  FROM
    (
    SELECT
      CASE WHEN atttypid=16 THEN ''DECODE(''||column_name||'',true,1,0)'' ELSE
 column_name END column_name
```

```
      ,atttypid
      ,''CAST(DECODE(number_of_items * sum_square_''||rn||'' - square_sum_''||
rn||'',0,null,(number_of_items*sum_target_''||rn||'' - sum_target * sum_''||rn||
        '')/SQRT((number_of_items * sum_square_target - square_sum_target) *
 (number_of_items * sum_square_''||rn||
        '' - square_sum_''||rn||''))) AS numeric(5,2)) ''||column_name
 outer_calculation
      ,''sum(''||column_name||'') sum_''||rn||'',''||
            ''SUM(trip_count*''||column_name||'') sum_target_''||rn||'',''||
            ''SUM(POW(''||column_name||'',2)) sum_square_''||rn||'',''||
            ''POW(SUM(''||column_name||''),2) square_sum_''||rn inner_calculation
    FROM
      (
      SELECT
        row_number() OVER (order by a.attnum) rn
        ,a.attname::VARCHAR column_name
        ,a.atttypid
      FROM pg_namespace AS n
        INNER JOIN pg_class AS c ON n.oid = c.relnamespace
        INNER JOIN pg_attribute AS a ON c.oid = a.attrelid
      WHERE a.attnum > 0
        AND n.nspname = '''||v_source_schema_name||'''
        AND c.relname = '''||v_source_table_name||'''
        AND a.atttypid IN (16,20,21,23,700,701,1700)
      )
    )
)';
  EXECUTE v_sql INTO v_generated_sql;
  EXECUTE  v_generated_sql;
END;
$$ LANGUAGE plpgsql;
```

La siguiente consulta muestra la correlación entre la columna objetivo, `trip_count`, y otros atributos numéricos de nuestro conjunto de datos.

```
call sp_correlation(
    'public',
    'trip_data',
    'trip_count',
    'tmp_corr_table'
);

SELECT
```

```
    *
FROM
    tmp_corr_table;
```

A continuación se muestra la salida de la operación `sp_correlation` anterior.

```
+-----------+----------+---------+-----------+----------+-------------
+-----------------+--------------+--------+-----------------+-----------
+---------------+
| trip_count | trip_hour | trip_day | trip_month | trip_year | trip_quarter
 | trip_month_week | trip_week_day | temp_c | precip_amount_mm | is_holiday |
 serial_number |
+-----------+----------+---------+-----------+----------+-------------
+-----------------+--------------+--------+-----------------+-----------
+---------------+
|          1 |      0.32 |     0.01 |       0.18 |      0.12 |         0.18 |
     0 |          0.02 |   0.53 |            -0.07 |      -0.13 |             0 |
+-----------+----------+---------+-----------+----------+-------------
+-----------------+--------------+--------+-----------------+-----------
+---------------+
```

## Paso 2: Crear el modelo de machine learning

1. La siguiente consulta divide los datos en un conjunto de entrenamiento y un conjunto de validación al designar el 80 % del conjunto de datos para el entrenamiento y el 20 % para la validación. El conjunto de entrenamiento es la entrada del modelo de ML para identificar el mejor algoritmo posible para el modelo. Una vez creado el modelo, se utiliza el conjunto de validación para validar la precisión del modelo.

   ```
   CREATE TABLE training_data AS
   SELECT
       trip_count,
       trip_hour,
       trip_day,
       trip_month,
       trip_year,
       trip_quarter,
       trip_month_week,
       trip_week_day,
       temp_c,
       precip_amount_mm,
       is_holiday
   ```

```
FROM
    trip_data
WHERE
    serial_number > (
        SELECT
            COUNT(1) * 0.2
        FROM
            trip_data
    );

CREATE TABLE validation_data AS
SELECT
    trip_count,
    trip_hour,
    trip_day,
    trip_month,
    trip_year,
    trip_quarter,
    trip_month_week,
    trip_week_day,
    temp_c,
    precip_amount_mm,
    is_holiday,
    trip_time
FROM
    trip_data
WHERE
    serial_number <= (
        SELECT
            COUNT(1) * 0.2
        FROM
            trip_data
    );
```

2. La siguiente consulta crea un modelo de regresión para predecir el valor de `trip_count` para cualquier fecha y hora de entrada. En el siguiente ejemplo, reemplace amzn-s3-demo-bucket por el bucket de S3 propio.

```
CREATE MODEL predict_rental_count
FROM
    training_data TARGET trip_count FUNCTION predict_rental_count
    IAM_ROLE default
    PROBLEM_TYPE regression
```

```
    OBJECTIVE 'mse'
    SETTINGS (
        s3_bucket 'amzn-s3-demo-bucket',
        s3_garbage_collect off,
        max_runtime 5000
    );
```

## Paso 3: Validar el modelo

1. Utilice la siguiente consulta para generar aspectos del modelo y buscar la métrica del error cuadrado medio raíz en la salida. El error cuadrado medio es una métrica de precisión típica para los problemas de regresión.

```
show model predict_rental_count;
```

2. Ejecute las siguientes consultas de predicción con los datos de validación para comparar el recuento de viajes previsto con el recuento de viajes real.

```
SELECT
    trip_time,
    actual_count,
    predicted_count,
    (actual_count - predicted_count) difference
FROM
    (
        SELECT
            trip_time,
            trip_count AS actual_count,
            PREDICT_RENTAL_COUNT (
                trip_hour,
                trip_day,
                trip_month,
                trip_year,
                trip_quarter,
                trip_month_week,
                trip_week_day,
                temp_c,
                precip_amount_mm,
                is_holiday
            ) predicted_count
        FROM
```

```
            validation_data
    )
LIMIT
    5;
```

3. La siguiente consulta calcula el error cuadrado medio y el error cuadrado medio raíz basándose en los datos de validación. El error cuadrado medio y el error cuadrado medio raíz se utilizan para medir la distancia entre el destino numérico predicho y la respuesta numérica real. Un buen modelo tiene una puntuación baja en ambas métricas. La siguiente consulta devuelve el valor de ambas métricas.

```
SELECT
    ROUND(
        AVG(POWER((actual_count - predicted_count), 2)),
        2
    ) mse,
    ROUND(
        SQRT(AVG(POWER((actual_count - predicted_count), 2))),
        2
    ) rmse
FROM
    (
        SELECT
            trip_time,
            trip_count AS actual_count,
            PREDICT_RENTAL_COUNT (
                trip_hour,
                trip_day,
                trip_month,
                trip_year,
                trip_quarter,
                trip_month_week,
                trip_week_day,
                temp_c,
                precip_amount_mm,
                is_holiday
            ) predicted_count
        FROM
            validation_data
    );
```

4. La siguiente consulta calcula el porcentaje de error en el recuento de viajes para cada tiempo de viaje el 01/01/2017. La consulta ordena los tiempos de viaje desde el tiene el error porcentual más bajo hasta el tiempo con el error porcentual más alto.

```
SELECT
    trip_time,
    CAST(ABS(((actual_count - predicted_count) / actual_count)) * 100 AS DECIMAL
 (7,2)) AS pct_error
FROM
    (
        SELECT
            trip_time,
            trip_count AS actual_count,
            PREDICT_RENTAL_COUNT (
                trip_hour,
                trip_day,
                trip_month,
                trip_year,
                trip_quarter,
                trip_month_week,
                trip_week_day,
                temp_c,
                precip_amount_mm,
                is_holiday
            ) predicted_count
        FROM
            validation_data
    )
WHERE
   trip_time LIKE '2017-01-01 %%:%%:%%'
ORDER BY
   2 ASC;
```

## Temas relacionados de

Para obtener más información sobre Amazon Redshift ML, consulte la siguiente documentación:

- Costos de utilizar Amazon Redshift ML
- Operación CREATE MODEL
- Función EXPLAIN_MODEL

Para obtener más información sobre machine learning, consulte la siguiente documentación:

- Información general sobre machine learning
- Machine learning para principiantes y expertos
- ¿Qué es la equidad y la explicabilidad del modelo para las predicciones de machine learning?

## Tutorial: Creación de modelos de regresión con aprendizaje lineal

En este tutorial, creará un modelo de aprendizaje lineal con datos de Amazon S3 y ejecutará consultas de predicción con el modelo mediante Amazon Redshift ML. El algoritmo de aprendizaje lineal de SageMaker AI resuelve problemas de regresión o clasificación multiclase. Para obtener más información sobre los problemas de regresión y clasificación multiclase, consulte Problem types for the machine learning paradigms en la Guía para desarrolladores de Amazon SageMaker AI. En este tutorial, resolverá un problema de regresión. El algoritmo de aprendizaje lineal entrena muchos modelos en paralelo y determina automáticamente el modelo más optimizado. Utilice la operación CREATE MODEL en Amazon Redshift, que crea el modelo de aprendizaje lineal mediante SageMaker AI y envía una función de predicción a Amazon Redshift. Para obtener más información sobre el algoritmo de aprendizaje lineal, consulte Algoritmo de aprendizaje lineal en la Guía para desarrolladores de Amazon SageMaker AI.

Se puede utilizar un comando CREATE MODEL para exportar datos de entrenamiento, entrenar un modelo, importar ese modelo y preparar una función de predicción de Amazon Redshift. Utilice la operación CREATE MODEL para especificar los datos de entrenamiento en forma de tabla o con la instrucción SELECT.

Los modelos de aprendizaje lineal optimizan los objetivos continuos o los objetivos discretos. Los objetivos continuos se utilizan para la regresión, mientras que las variables discretas se utilizan para la clasificación. Algunos métodos proporcionan una solución solo para objetivos continuos, como el método de regresión. El algoritmo de aprendizaje lineal proporciona un aumento de la velocidad comparado con las técnicas de optimización de hiperparámetros ingenua, como la técnica Naive Bayes. La técnica de optimización ingenua supone que cada variable de entrada es independiente. Para utilizar el algoritmo de aprendizaje lineal, debe proporcionar columnas que representan las dimensiones de las entradas y filas que representan las observaciones. Para obtener más información sobre el algoritmo de aprendizaje lineal, consulte Algoritmo de aprendizaje lineal en la Guía para desarrolladores de Amazon SageMaker AI.

En este tutorial, vamos a crear un modelo de aprendizaje lineal que predice la edad del abulón. Utilice el comando CREATE MODEL en el conjunto de datos del abulón para determinar la relación entre las medidas físicas del abulón. Luego, utilice el modelo para determinar la edad del abulón.

## Ejemplos de casos de uso

Puede resolver otros problemas de regresión con el aprendizaje lineal y Amazon Redshift ML, como predecir el precio de una casa. También puede usar Redshift ML para predecir la cantidad de personas que usarán el servicio de alquiler de bicicletas de una ciudad.

Tareas

- Requisitos previos
- Paso 1: Cargar los datos desde Amazon S3 en Amazon Redshift
- Paso 2: Crear el modelo de machine learning
- Paso 3: Validar el modelo

## Requisitos previos

Para completar este tutorial, debe realizar el procedimiento de configuración administrativa de Amazon Redshift ML.

## Paso 1: Cargar los datos desde Amazon S3 en Amazon Redshift

Use el editor de consultas de Amazon Redshift v2 para ejecutar las siguientes consultas. Estas consultas cargan los datos de muestra en Redshift y los dividen en un conjunto de entrenamiento y un conjunto de validación.

1. En la siguiente consulta, se crea la tabla `abalone_dataset`.

```
CREATE TABLE abalone_dataset (
    id INT IDENTITY(1, 1),
    Sex CHAR(1),
    Length float,
    Diameter float,
    Height float,
    Whole float,
    Shucked float,
    Viscera float,
    Shell float,
```

```
    Rings integer
);
```

2. La siguiente consulta copia los datos de muestra del conjunto de datos de abulón en Amazon S3 en la tabla `abalone_dataset` que creó anteriormente en Amazon Redshift.

```
COPY abalone_dataset
FROM
    's3://redshift-ml-multiclass/abalone.csv' REGION 'us-east-1' IAM_ROLE default CSV
 IGNOREHEADER 1 NULL AS 'NULL';
```

3. Al dividir manualmente los datos, podrá verificar la precisión del modelo asignando un conjunto de predicciones adicional. La siguiente consulta divide los datos en dos conjuntos. La tabla `abalone_training` es para el entrenamiento y la tabla `abalone_validationl` es para validación.

```
CREATE TABLE abalone_training as
SELECT
    *
FROM
    abalone_dataset
WHERE
    mod(id, 10) < 8;

CREATE TABLE abalone_validation as
SELECT
    *
FROM
    abalone_dataset
WHERE
    mod(id, 10) >= 8;
```

## Paso 2: Crear el modelo de machine learning

En este paso, utilizará la instrucción CREATE MODEL para crear el modelo de machine learning con el algoritmo de aprendizaje lineal.

La siguiente consulta crea el modelo de aprendizaje lineal con la operación CREATE MODEL mediante el bucket de S3. Reemplace amzn-s3-demo-bucket por el bucket de S3 propio.

```
CREATE MODEL model_abalone_ring_prediction
```

```
FROM
    (
        SELECT
            Sex,
            Length,
            Diameter,
            Height,
            Whole,
            Shucked,
            Viscera,
            Shell,
            Rings AS target_label
        FROM
            abalone_training
    ) TARGET target_label FUNCTION f_abalone_ring_prediction IAM_ROLE default
 MODEL_TYPE LINEAR_LEARNER PROBLEM_TYPE REGRESSION OBJECTIVE 'MSE' SETTINGS (
        S3_BUCKET 'amzn-s3-demo-bucket',
        MAX_RUNTIME 15000
    );
```

Comprobar el estado del entrenamiento del modelo (opcional)

Puede utilizar el comando SHOW MODEL para saber cuándo está listo el modelo.

Utilice la siguiente consulta para supervisar el estado del entrenamiento del modelo.

```
SHOW MODEL model_abalone_ring_prediction;
```

Cuando el modelo esté listo, la salida de la operación anterior tendrá un aspecto semejante al del siguiente ejemplo. Tenga en cuenta que la salida proporciona la métrica `validation:mse`, que es el error cuadrado medio. Utilice el error cuadrado medio para validar la precisión del modelo en el paso siguiente.

```
+--------------------------
+-----------------------------------------------------------------------------------------------
+
|        Model Name        |
 model_abalone_ring_prediction                                    |
+--------------------------
+-----------------------------------------------------------------------------------------------
+
| Schema Name              | public
                                        |
```

```
| Owner                     | awsuser
                                      |
| Creation Time             | Thu, 30.06.2022 18:00:10
                                      |
| Model State               | READY
                                      |
| validation:mse            |
                                  4.168633 |
| Estimated Cost            |
                                  4.291608 |
|                           |
                                      |
| TRAINING DATA:            |
                                      |
| Query                     | SELECT SEX , LENGTH , DIAMETER , HEIGHT , WHOLE ,
 SHUCKED , VISCERA , SHELL, RINGS AS TARGET_LABEL |
|                           | FROM ABALONE_TRAINING
                                      |
| Target Column             | TARGET_LABEL
                                      |
|                           |
                                      |
| PARAMETERS:               |
                                      |
| Model Type                | linear_learner
                                      |
| Problem Type              | Regression
                                      |
| Objective                 | MSE
                                      |
| AutoML Job Name           | redshiftml-20220630180010947843
                                      |
| Function Name             | f_abalone_ring_prediction
                                      |
| Function Parameters       | sex length diameter height whole shucked viscera shell
                                      |
| Function Parameter Types | bpchar float8 float8 float8 float8 float8 float8 float8
                                      |
| IAM Role                  | default-aws-iam-role
                                      |
| S3 Bucket                 | amzn-s3-demo-bucket
                                      |
| Max Runtime               |
                                 15000 |
```

```
+-------------------------
+-------------------------------------------------------------------------------------------------------
+
```

## Paso 3: Validar el modelo

1. La siguiente consulta de predicción valida la precisión del modelo en el conjunto de datos `abalone_validation`calculando el error cuadrado medio y el error cuadrado medio raíz.

```
SELECT
    ROUND(AVG(POWER((tgt_label - predicted), 2)), 2) mse,
    ROUND(SQRT(AVG(POWER((tgt_label - predicted), 2))), 2) rmse
FROM
    (
        SELECT
            Sex,
            Length,
            Diameter,
            Height,
            Whole,
            Shucked,
            Viscera,
            Shell,
            Rings AS tgt_label,
            f_abalone_ring_prediction(
                Sex,
                Length,
                Diameter,
                Height,
                Whole,
                Shucked,
                Viscera,
                Shell
            ) AS predicted,
            CASE
                WHEN tgt_label = predicted then 1
                ELSE 0
            END AS match,
            CASE
                WHEN tgt_label <> predicted then 1
                ELSE 0
            END AS nonmatch
        FROM
```

```
            abalone_validation
    ) t1;
```

La salida de la consulta anterior tendrá un aspecto semejante al de este ejemplo. El valor de la métrica del error cuadrado medio debe ser similar al valor de la métrica `validation:mse` que se muestra en la salida de la operación SHOW MODEL.

```
+-----+--------------------+
| mse |        rmse        |
+-----+--------------------+
| 5.1 | 2.2600000000000002 |
+-----+--------------------+
```

2. Utilice la siguiente consulta para ejecutar la operación EXPLAIN_MODEL en la función de predicción. La operación devolverá un informe de explicabilidad del modelo. Para obtener más información sobre la operación EXPLAIN_MODEL, consulte Función EXPLAIN_MODEL en la Guía para desarrolladores de bases de datos de Amazon Redshift.

```
SELECT
    EXPLAIN_MODEL ('model_abalone_ring_prediction');
```

La siguiente información es un ejemplo del informe de explicabilidad del modelo generado por la operación EXPLAIN_MODEL anterior. Los valores de cada una de las entradas son valores de Shapley. Los valores de Shapley representan el efecto que tiene cada entrada en la predicción de su modelo, donde las entradas de mayor valor tienen más impacto en la predicción. En este ejemplo, las entradas de mayor valor tienen más impacto en la predicción de la edad del abulón.

```
{
    "explanations": {
        "kernel_shap": {
            "label0": {
                "expected_value" :10.290688514709473,
                "global_shap_values": {
                    "diameter" :0.6856910187882492,
                    "height" :0.4415323937124035,
                    "length" :0.21507476107609084,
                    "sex" :0.448611774505744,
                    "shell" :1.70426496893776,
                    "shucked" :2.1181392924386994,
                    "viscera" :0.342220754059912,
```

```
                    "whole" :0.6711906974084011
                }
            }
        }
    },
    "version" :"1.0"
};
```

3. Utilice la siguiente consulta para calcular el porcentaje de predicciones correctas que el modelo hace sobre los abulones que aún no están maduros. Los abulones inmaduros tienen 10 anillos o menos, y una predicción correcta es precisa dentro de un anillo del número real de anillos.

```
SELECT
    TRUNC(
        SUM(
            CASE
                WHEN ROUND(
                    f_abalone_ring_prediction(
                        Sex,
                        Length,
                        Diameter,
                        Height,
                        Whole,
                        Shucked,
                        Viscera,
                        Shell
                    ),
                    0
                ) BETWEEN Rings - 1
                AND Rings + 1 THEN 1
                ELSE 0
            END
        ) / CAST(COUNT(SHELL) AS FLOAT),
        4
    ) AS prediction_pct
FROM
    abalone_validation
WHERE
    Rings <= 10;
```

### Temas relacionados de

Para obtener más información sobre Amazon Redshift ML, consulte la siguiente documentación:

- Costos de utilizar Amazon Redshift ML
- Operación CREATE MODEL
- Función EXPLAIN_MODEL

Para obtener más información sobre machine learning, consulte la siguiente documentación:

- Información general sobre machine learning
- Machine learning para principiantes y expertos
- ¿Qué es la equidad y la explicabilidad del modelo para las predicciones de machine learning?

## Tutorial: Creación de modelos de clasificación multiclase con aprendizaje lineal

En este tutorial, creará un modelo de aprendizaje lineal con datos de Amazon S3 y, a continuación, ejecutará consultas de predicción con el modelo mediante Amazon Redshift ML. El algoritmo de aprendizaje lineal de SageMaker AI resuelve problemas de regresión o clasificación. Para obtener más información sobre los problemas de regresión y clasificación multiclase, consulte Problem types for the machine learning paradigms en la Guía para desarrolladores de Amazon SageMaker AI.
En este tutorial, solucionará un problema de clasificación multiclase. El algoritmo de aprendizaje lineal entrena muchos modelos en paralelo y determina automáticamente el modelo más optimizado. Utilice la operación CREATE MODEL en Amazon Redshift, que crea el modelo de aprendizaje lineal mediante SageMaker AI y envía la función de predicción a Amazon Redshift. Para obtener más información sobre el algoritmo de aprendizaje lineal, consulte Algoritmo de aprendizaje lineal en la Guía para desarrolladores de Amazon SageMaker AI.

Se puede utilizar un comando CREATE MODEL para exportar datos de entrenamiento, entrenar un modelo, importar ese modelo y preparar una función de predicción de Amazon Redshift. Utilice la operación CREATE MODEL para especificar los datos de entrenamiento en forma de tabla o con la instrucción SELECT.

Los modelos de aprendizaje lineal optimizan los objetivos continuos o los objetivos discretos. Los objetivos continuos se utilizan para la regresión, mientras que las variables discretas se utilizan para la clasificación. Algunos métodos proporcionan una solución solo para objetivos continuos, como

un método de regresión. El algoritmo de aprendizaje lineal proporciona un aumento de la velocidad comparado con las técnicas de optimización de hiperparámetros ingenua, como la técnica Naive Bayes. La técnica de optimización ingenua supone que cada variable de entrada es independiente. El algoritmo de aprendizaje lineal entrena muchos modelos en paralelo y selecciona el modelo más optimizado. Un algoritmo similar es XGBoost, que combina estimaciones de un conjunto de modelos más simples y más débiles para hacer predicciones. Para obtener más información sobre XGBoost, consulte Algoritmo XGBoost en la Guía para desarrolladores de Amazon SageMaker AI.

Para utilizar el algoritmo de aprendizaje lineal, debe proporcionar columnas que representan las dimensiones de las entradas y filas que representan las observaciones. Para obtener más información sobre el algoritmo de aprendizaje lineal, consulte Algoritmo de aprendizaje lineal en la Guía para desarrolladores de Amazon SageMaker AI.

En este tutorial, creará un modelo de aprendizaje lineal que predice los tipos de cubiertas para un área determinada. Utilice el comando CREATE MODEL en el conjunto de datos Covertype de UCI Machine Learning Repository. A continuación, utilice la función de predicción creada por el comando para determinar los tipos de cubierta en un área natural. Un tipo de cubierta forestal suele ser un tipo de árbol. Las entradas que utilizará Redshift ML para crear el modelo incluyen el tipo de suelo, la distancia a las carreteras y la designación de áreas naturales. Para obtener más información sobre el conjunto de datos, consulte el conjunto de datos Covertype de UCI Machine Learning Repository.

## Ejemplos de casos de uso

Puede resolver otros problemas de clasificación multiclase con el aprendizaje lineal con Amazon Redshift ML, como predecir la especie de una planta a partir de una imagen. También puede predecir la cantidad de un producto que comprará un cliente.

Tareas

- Requisitos previos
- Paso 1: Cargar los datos desde Amazon S3 en Amazon Redshift
- Paso 2: Crear el modelo de machine learning
- Paso 3: Validar el modelo

## Requisitos previos

Para completar este tutorial, debe realizar el procedimiento de configuración administrativa de Amazon Redshift ML.

## Paso 1: Cargar los datos desde Amazon S3 en Amazon Redshift

Use el editor de consultas de Amazon Redshift v2 para ejecutar las siguientes consultas. Estas consultas cargan los datos de muestra en Redshift y los dividen en un conjunto de entrenamiento y un conjunto de validación.

1. En la siguiente consulta, se crea la tabla `covertype_data`.

```
CREATE TABLE public.covertype_data (
    elevation bigint ENCODE az64,
    aspect bigint ENCODE az64,
    slope bigint ENCODE az64,
    horizontal_distance_to_hydrology bigint ENCODE az64,
    vertical_distance_to_hydrology bigint ENCODE az64,
    horizontal_distance_to_roadways bigint ENCODE az64,
    hillshade_9am bigint ENCODE az64,
    hillshade_noon bigint ENCODE az64,
    hillshade_3pm bigint ENCODE az64,
    horizontal_distance_to_fire_points bigint ENCODE az64,
    wilderness_area1 bigint ENCODE az64,
    wilderness_area2 bigint ENCODE az64,
    wilderness_area3 bigint ENCODE az64,
    wilderness_area4 bigint ENCODE az64,
    soil_type1 bigint ENCODE az64,
    soil_type2 bigint ENCODE az64,
    soil_type3 bigint ENCODE az64,
    soil_type4 bigint ENCODE az64,
    soil_type5 bigint ENCODE az64,
    soil_type6 bigint ENCODE az64,
    soil_type7 bigint ENCODE az64,
    soil_type8 bigint ENCODE az64,
    soil_type9 bigint ENCODE az64,
    soil_type10 bigint ENCODE az64,
    soil_type11 bigint ENCODE az64,
    soil_type12 bigint ENCODE az64,
    soil_type13 bigint ENCODE az64,
    soil_type14 bigint ENCODE az64,
    soil_type15 bigint ENCODE az64,
    soil_type16 bigint ENCODE az64,
    soil_type17 bigint ENCODE az64,
    soil_type18 bigint ENCODE az64,
    soil_type19 bigint ENCODE az64,
    soil_type20 bigint ENCODE az64,
```

```
    soil_type21 bigint ENCODE az64,
    soil_type22 bigint ENCODE az64,
    soil_type23 bigint ENCODE az64,
    soil_type24 bigint ENCODE az64,
    soil_type25 bigint ENCODE az64,
    soil_type26 bigint ENCODE az64,
    soil_type27 bigint ENCODE az64,
    soil_type28 bigint ENCODE az64,
    soil_type29 bigint ENCODE az64,
    soil_type30 bigint ENCODE az64,
    soil_type31 bigint ENCODE az64,
    soil_type32 bigint ENCODE az64,
    soil_type33 bigint ENCODE az64,
    soil_type34 bigint ENCODE az64,
    soil_type35 bigint ENCODE az64,
    soil_type36 bigint ENCODE az64,
    soil_type37 bigint ENCODE az64,
    soil_type38 bigint ENCODE az64,
    soil_type39 bigint ENCODE az64,
    soil_type40 bigint ENCODE az64,
    cover_type bigint ENCODE az64
) DISTSTYLE AUTO;
```

2. La siguiente consulta copia los datos de muestra del conjunto de datos Covertype en Amazon S3 en la tabla `covertype_data` que creó anteriormente en Amazon Redshift.

```
COPY public.covertype_data
FROM
    's3://redshift-ml-multiclass/covtype.data.gz' IAM_ROLE DEFAULT gzip DELIMITER ','
 REGION 'us-east-1';
```

3. Al dividir manualmente los datos, podrá verificar la precisión del modelo asignando un conjunto de pruebas adicional. La siguiente consulta divide los datos en tres conjuntos. La tabla `covertype_training` es para el entrenamiento, la tabla `covertype_validation` es para la validación y la tabla `covertype_test` es para probar su modelo. Utilizará el conjunto de entrenamiento para entrenar el modelo y el conjunto de validación para validar el desarrollo del modelo. A continuación, utilizará el conjunto de pruebas para probar el rendimiento del modelo y comprobar si el modelo está sobreajustado o insuficientemente ajustado en el conjunto de datos.

```
CREATE TABLE public.covertype_data_prep AS
SELECT
    a.*,
```

```
    CAST (random() * 100 AS int) AS data_group_id
FROM
    public.covertype_data a;

--training dataset
 CREATE TABLE public.covertype_training as
SELECT
    *
FROM
    public.covertype_data_prep
WHERE
    data_group_id < 80;

--validation dataset
 CREATE TABLE public.covertype_validation AS
SELECT
    *
FROM
    public.covertype_data_prep
WHERE
    data_group_id BETWEEN 80
    AND 89;

--test dataset
 CREATE TABLE public.covertype_test AS
SELECT
    *
FROM
    public.covertype_data_prep
WHERE
    data_group_id > 89;
```

## Paso 2: Crear el modelo de machine learning

En este paso, utilizará la instrucción CREATE MODEL para crear el modelo de machine learning con el algoritmo de aprendizaje lineal.

La siguiente consulta crea el modelo de aprendizaje lineal con la operación CREATE MODEL mediante el bucket de S3. Reemplace amzn-s3-demo-bucket por el bucket de S3 propio.

```
CREATE MODEL forest_cover_type_model
FROM
```

```
    (
        SELECT
            Elevation,
            Aspect,
            Slope,
            Horizontal_distance_to_hydrology,
            Vertical_distance_to_hydrology,
            Horizontal_distance_to_roadways,
            HIllshade_9am,
            Hillshade_noon,
            Hillshade_3pm,
            Horizontal_Distance_To_Fire_Points,
            Wilderness_Area1,
            Wilderness_Area2,
            Wilderness_Area3,
            Wilderness_Area4,
            soil_type1,
            Soil_Type2,
            Soil_Type3,
            Soil_Type4,
            Soil_Type5,
            Soil_Type6,
            Soil_Type7,
            Soil_Type8,
            Soil_Type9,
            Soil_Type10,
            Soil_Type11,
            Soil_Type12,
            Soil_Type13,
            Soil_Type14,
            Soil_Type15,
            Soil_Type16,
            Soil_Type17,
            Soil_Type18,
            Soil_Type19,
            Soil_Type20,
            Soil_Type21,
            Soil_Type22,
            Soil_Type23,
            Soil_Type24,
            Soil_Type25,
            Soil_Type26,
            Soil_Type27,
            Soil_Type28,
```

```
            Soil_Type29,
            Soil_Type30,
            Soil_Type31,
            Soil_Type32,
            Soil_Type33,
            Soil_Type34,
            Soil_Type36,
            Soil_Type37,
            Soil_Type38,
            Soil_Type39,
            Soil_Type40,
            Cover_type
        from
            public.covertype_training
    ) TARGET cover_type FUNCTION predict_cover_type IAM_ROLE default MODEL_TYPE
 LINEAR_LEARNER PROBLEM_TYPE MULTICLASS_CLASSIFICATION OBJECTIVE 'Accuracy' SETTINGS (
        S3_BUCKET 'amzn-s3-demo-bucket',
        S3_GARBAGE_COLLECT OFF,
        MAX_RUNTIME 15000
    );
```

Comprobar el estado del entrenamiento del modelo (opcional)

Puede utilizar el comando SHOW MODEL para saber cuándo está listo el modelo.

Utilice la siguiente consulta para supervisar el estado del entrenamiento del modelo.

```
SHOW MODEL forest_cover_type_model;
```

Cuando el modelo esté listo, la salida de la operación anterior tendrá un aspecto semejante al del siguiente ejemplo. Tenga en cuenta que en la salida se proporciona la métrica `validation:multiclass_accuracy`, que puede ver en el lado derecho del siguiente ejemplo. La precisión multiclase mide el porcentaje de puntos de datos que el modelo clasifica correctamente. Utilizará la precisión multiclase para validar la precisión del modelo en el siguiente paso.

```
+--------------------------------
+-------------------------------------------------------------------------------------------------------------------------------
+
|             Key               |
```

```
 Value


                                                                |
+--------------------------------
+-------------------------------------------------------------------------------------------------------
+
| Model Name                     | forest_cover_type_model


                                                                |
| Schema Name                    | public


                                                                |
| Owner                          | awsuser


                                                                |
| Creation Time                  | Tue, 12.07.2022 20:24:32
```

```
                                                    |
| Model State                    | READY

                                                    |
| validation:multiclass_accuracy |

                                             0.724952 |
| Estimated Cost                 |

                                             5.341750 |
|                                |
```

```
                                                    |
| TRAINING DATA:                  |

                                                    |
| Query                           | SELECT ELEVATION, ASPECT, SLOPE,
 HORIZONTAL_DISTANCE_TO_HYDROLOGY, VERTICAL_DISTANCE_TO_HYDROLOGY,
 HORIZONTAL_DISTANCE_TO_ROADWAYS, HILLSHADE_9AM, HILLSHADE_NOON, HILLSHADE_3PM ,
 HORIZONTAL_DISTANCE_TO_FIRE_POINTS, WILDERNESS_AREA1, WILDERNESS_AREA2,
 WILDERNESS_AREA3, WILDERNESS_AREA4, SOIL_TYPE1, SOIL_TYPE2, SOIL_TYPE3, SOIL_TYPE4,
 SOIL_TYPE5, SOIL_TYPE6, SOIL_TYPE7, SOIL_TYPE8, SOIL_TYPE9, SOIL_TYPE10 , SOIL_TYPE11,
 SOIL_TYPE12 , SOIL_TYPE13 , SOIL_TYPE14, SOIL_TYPE15, SOIL_TYPE16, SOIL_TYPE17,
 SOIL_TYPE18, SOIL_TYPE19, SOIL_TYPE20, SOIL_TYPE21, SOIL_TYPE22, SOIL_TYPE23,
 SOIL_TYPE24, SOIL_TYPE25, SOIL_TYPE26, SOIL_TYPE27, SOIL_TYPE28, SOIL_TYPE29,
 SOIL_TYPE30, SOIL_TYPE31, SOIL_TYPE32, SOIL_TYPE33, SOIL_TYPE34, SOIL_TYPE36,
 SOIL_TYPE37, SOIL_TYPE38, SOIL_TYPE39, SOIL_TYPE40, COVER_TYPE |
|                                 | FROM PUBLIC.COVERTYPE_TRAINING

                                                    |
| Target Column                   | COVER_TYPE

                                                    |
```

```
|                                |                                                
```

```
                                                    |
| AutoML Job Name                | redshiftml-20220712202432187659

                                                    |
| Function Name                  | predict_cover_type

                                                    |
| Function Parameters            | elevation aspect slope
 horizontal_distance_to_hydrology vertical_distance_to_hydrology
 horizontal_distance_to_roadways hillshade_9am hillshade_noon hillshade_3pm
 horizontal_distance_to_fire_points wilderness_area1 wilderness_area2 wilderness_area3
 wilderness_area4 soil_type1 soil_type2 soil_type3 soil_type4 soil_type5 soil_type6
 soil_type7 soil_type8 soil_type9 soil_type10 soil_type11 soil_type12 soil_type13
 soil_type14 soil_type15 soil_type16 soil_type17 soil_type18 soil_type19 soil_type20
 soil_type21 soil_type22 soil_type23 soil_type24 soil_type25 soil_type26 soil_type27
 soil_type28 soil_type29 soil_type30 soil_type31 soil_type32 soil_type33 soil_type34
 soil_type36 soil_type37 soil_type38 soil_type39 soil_type40
                                                   |
| Function Parameter Types       | int8 int8 int8 int8 int8 int8 int8 int8 int8 int8
 int8 int8 int8 int8 int8 int8 int8 int8 int8 int8 int8 int8 int8 int8 int8 int8 int8
 int8 int8 int8 int8 int8 int8 int8 int8 int8 int8 int8 int8 int8 int8 int8 int8 int8
 int8 int8 int8 int8 int8 int8 int8 int8 int8
```

```
                                                    |
| IAM Role                        | default-aws-iam-role

                                                    |
| S3 Bucket                       | amzn-s3-demo-bucket

                                                    |
| Max Runtime                     |

                                                15000 |
+---------------------------------
+-----------------------------------------------------------------------------------------------
+
```

## Paso 3: Validar el modelo

1. La siguiente consulta de predicción valida la precisión del modelo en el conjunto de datos `covertype_validation` calculando la precisión multiclase. La precisión multiclase es el porcentaje de las predicciones del modelo que son correctas.

```
SELECT
    CAST(sum(t1.match) AS decimal(7, 2)) AS predicted_matches,
    CAST(sum(t1.nonmatch) AS decimal(7, 2)) AS predicted_non_matches,
    CAST(sum(t1.match + t1.nonmatch) AS decimal(7, 2)) AS total_predictions,
    predicted_matches / total_predictions AS pct_accuracy
FROM
    (
        SELECT
            Elevation,
            Aspect,
            Slope,
            Horizontal_distance_to_hydrology,
            Vertical_distance_to_hydrology,
            Horizontal_distance_to_roadways,
            HIllshade_9am,
            Hillshade_noon,
            Hillshade_3pm,
            Horizontal_Distance_To_Fire_Points,
            Wilderness_Area1,
            Wilderness_Area2,
            Wilderness_Area3,
            Wilderness_Area4,
            soil_type1,
            Soil_Type2,
            Soil_Type3,
            Soil_Type4,
            Soil_Type5,
            Soil_Type6,
            Soil_Type7,
            Soil_Type8,
            Soil_Type9,
            Soil_Type10,
            Soil_Type11,
            Soil_Type12,
            Soil_Type13,
            Soil_Type14,
            Soil_Type15,
            Soil_Type16,
            Soil_Type17,
            Soil_Type18,
            Soil_Type19,
            Soil_Type20,
            Soil_Type21,
```

```
            Soil_Type22,
            Soil_Type23,
            Soil_Type24,
            Soil_Type25,
            Soil_Type26,
            Soil_Type27,
            Soil_Type28,
            Soil_Type29,
            Soil_Type30,
            Soil_Type31,
            Soil_Type32,
            Soil_Type33,
            Soil_Type34,
            Soil_Type36,
            Soil_Type37,
            Soil_Type38,
            Soil_Type39,
            Soil_Type40,
            Cover_type AS actual_cover_type,
            predict_cover_type(
                Elevation,
                Aspect,
                Slope,
                Horizontal_distance_to_hydrology,
                Vertical_distance_to_hydrology,
                Horizontal_distance_to_roadways,
                HIllshade_9am,
                Hillshade_noon,
                Hillshade_3pm,
                Horizontal_Distance_To_Fire_Points,
                Wilderness_Area1,
                Wilderness_Area2,
                Wilderness_Area3,
                Wilderness_Area4,
                soil_type1,
                Soil_Type2,
                Soil_Type3,
                Soil_Type4,
                Soil_Type5,
                Soil_Type6,
                Soil_Type7,
                Soil_Type8,
                Soil_Type9,
                Soil_Type10,
```

```
            Soil_Type11,
            Soil_Type12,
            Soil_Type13,
            Soil_Type14,
            Soil_Type15,
            Soil_Type16,
            Soil_Type17,
            Soil_Type18,
            Soil_Type19,
            Soil_Type20,
            Soil_Type21,
            Soil_Type22,
            Soil_Type23,
            Soil_Type24,
            Soil_Type25,
            Soil_Type26,
            Soil_Type27,
            Soil_Type28,
            Soil_Type29,
            Soil_Type30,
            Soil_Type31,
            Soil_Type32,
            Soil_Type33,
            Soil_Type34,
            Soil_Type36,
            Soil_Type37,
            Soil_Type38,
            Soil_Type39,
            Soil_Type40
        ) AS predicted_cover_type,
        CASE
            WHEN actual_cover_type = predicted_cover_type THEN 1
            ELSE 0
        END AS match,
        CASE
            WHEN actual_cover_type <> predicted_cover_type THEN 1
            ELSE 0
        END AS nonmatch
    FROM
        public.covertype_validation
) t1;
```

La salida de la consulta anterior tendrá un aspecto semejante al de este ejemplo.
El valor de la métrica de precisión multiclase debe ser similar al valor de la métrica `validation:multiclass_accuracy` que se muestra en la salida de la operación SHOW MODEL.

```
+------------------+----------------------+------------------+-------------+
| predicted_matches | predicted_non_matches | total_predictions | pct_accuracy |
+------------------+----------------------+------------------+-------------+
|             41211 |                16324 |             57535 |   0.71627704 |
+------------------+----------------------+------------------+-------------+
```

2. La siguiente consulta predice el tipo de cubierta más común para `wilderness_area2`. Este conjunto de datos incluye cuatro áreas naturales y siete tipos de cubiertas. Un área natural puede tener varios tipos de cubiertas.

```
SELECT t1. predicted_cover_type, COUNT(*)
FROM
(
SELECT
   Elevation,
   Aspect,
   Slope,
   Horizontal_distance_to_hydrology,
   Vertical_distance_to_hydrology,
   Horizontal_distance_to_roadways,
   HIllshade_9am,
   Hillshade_noon,
   Hillshade_3pm ,
   Horizontal_Distance_To_Fire_Points,
   Wilderness_Area1,
   Wilderness_Area2,
   Wilderness_Area3,
   Wilderness_Area4,
   soil_type1,
   Soil_Type2,
   Soil_Type3,
   Soil_Type4,
   Soil_Type5,
   Soil_Type6,
   Soil_Type7,
   Soil_Type8,
```

```
    Soil_Type9,
    Soil_Type10 ,
    Soil_Type11,
    Soil_Type12 ,
    Soil_Type13 ,
    Soil_Type14,
    Soil_Type15,
    Soil_Type16,
    Soil_Type17,
    Soil_Type18,
    Soil_Type19,
    Soil_Type20,
    Soil_Type21,
    Soil_Type22,
    Soil_Type23,
    Soil_Type24,
    Soil_Type25,
    Soil_Type26,
    Soil_Type27,
    Soil_Type28,
    Soil_Type29,
    Soil_Type30,
    Soil_Type31,
    Soil_Type32,
    Soil_Type33,
    Soil_Type34,
    Soil_Type36,
    Soil_Type37,
    Soil_Type38,
    Soil_Type39,
    Soil_Type40,
    predict_cover_type( Elevation,
    Aspect,
    Slope,
    Horizontal_distance_to_hydrology,
    Vertical_distance_to_hydrology,
    Horizontal_distance_to_roadways,
    HIllshade_9am,
    Hillshade_noon,
    Hillshade_3pm ,
    Horizontal_Distance_To_Fire_Points,
    Wilderness_Area1,
    Wilderness_Area2,
    Wilderness_Area3,
```

```
   Wilderness_Area4,
   soil_type1,
   Soil_Type2,
   Soil_Type3,
   Soil_Type4,
   Soil_Type5,
   Soil_Type6,
   Soil_Type7,
   Soil_Type8,
   Soil_Type9,
   Soil_Type10,
   Soil_Type11,
   Soil_Type12,
   Soil_Type13,
   Soil_Type14,
   Soil_Type15,
   Soil_Type16,
   Soil_Type17,
   Soil_Type18,
   Soil_Type19,
   Soil_Type20,
   Soil_Type21,
   Soil_Type22,
   Soil_Type23,
   Soil_Type24,
   Soil_Type25,
   Soil_Type26,
   Soil_Type27,
   Soil_Type28,
   Soil_Type29,
   Soil_Type30,
   Soil_Type31,
   Soil_Type32,
   Soil_Type33,
   Soil_Type34,
   Soil_Type36,
   Soil_Type37,
   Soil_Type38,
   Soil_Type39,
   Soil_Type40) AS predicted_cover_type

FROM public.covertype_test
WHERE wilderness_area2 = 1)
t1
```

```
GROUP BY 1;
```

La salida de la operación anterior tendrá un aspecto semejante al del siguiente ejemplo. Esta salida significa que el modelo predijo que la mayor parte de la cubierta es de tipo 1, y que hay alguna cubierta de los tipos 2 y 7.

```
+----------------------+-------+
| predicted_cover_type | count |
+----------------------+-------+
|                    2 |   564 |
|                    7 |    97 |
|                    1 |  2309 |
+----------------------+-------+
```

3. La siguiente consulta muestra el tipo de cubierta más común en una sola área natural. La consulta muestra la cantidad de ese tipo de cubierta y el área natural del tipo de cubierta.

```
SELECT t1. predicted_cover_type, COUNT(*), wilderness_area
FROM
(
SELECT
   Elevation,
   Aspect,
   Slope,
   Horizontal_distance_to_hydrology,
   Vertical_distance_to_hydrology,
   Horizontal_distance_to_roadways,
   HIllshade_9am,
   Hillshade_noon,
   Hillshade_3pm ,
   Horizontal_Distance_To_Fire_Points,
   Wilderness_Area1,
   Wilderness_Area2,
   Wilderness_Area3,
   Wilderness_Area4,
   soil_type1,
   Soil_Type2,
   Soil_Type3,
   Soil_Type4,
   Soil_Type5,
   Soil_Type6,
   Soil_Type7,
```

```
    Soil_Type8,
    Soil_Type9,
    Soil_Type10 ,
    Soil_Type11,
    Soil_Type12 ,
    Soil_Type13 ,
    Soil_Type14,
    Soil_Type15,
    Soil_Type16,
    Soil_Type17,
    Soil_Type18,
    Soil_Type19,
    Soil_Type20,
    Soil_Type21,
    Soil_Type22,
    Soil_Type23,
    Soil_Type24,
    Soil_Type25,
    Soil_Type26,
    Soil_Type27,
    Soil_Type28,
    Soil_Type29,
    Soil_Type30,
    Soil_Type31,
    Soil_Type32,
    Soil_Type33,
    Soil_Type34,
    Soil_Type36,
    Soil_Type37,
    Soil_Type38,
    Soil_Type39,
    Soil_Type40,
    predict_cover_type( Elevation,
    Aspect,
    Slope,
    Horizontal_distance_to_hydrology,
    Vertical_distance_to_hydrology,
    Horizontal_distance_to_roadways,
    HIllshade_9am,
    Hillshade_noon,
    Hillshade_3pm ,
    Horizontal_Distance_To_Fire_Points,
    Wilderness_Area1,
    Wilderness_Area2,
```

```
    Wilderness_Area3,
    Wilderness_Area4,
    soil_type1,
    Soil_Type2,
    Soil_Type3,
    Soil_Type4,
    Soil_Type5,
    Soil_Type6,
    Soil_Type7,
    Soil_Type8,
    Soil_Type9,
    Soil_Type10,
    Soil_Type11,
    Soil_Type12,
    Soil_Type13,
    Soil_Type14,
    Soil_Type15,
    Soil_Type16,
    Soil_Type17,
    Soil_Type18,
    Soil_Type19,
    Soil_Type20,
    Soil_Type21,
    Soil_Type22,
    Soil_Type23,
    Soil_Type24,
    Soil_Type25,
    Soil_Type26,
    Soil_Type27,
    Soil_Type28,
    Soil_Type29,
    Soil_Type30,
    Soil_Type31,
    Soil_Type32,
    Soil_Type33,
    Soil_Type34,
    Soil_Type36,
    Soil_Type37,
    Soil_Type38,
    Soil_Type39,
    Soil_Type40) AS predicted_cover_type,
    CASE WHEN Wilderness_Area1 = 1 THEN 1
         WHEN Wilderness_Area2 = 1 THEN 2
         WHEN Wilderness_Area3 = 1 THEN 3
```

```
        WHEN Wilderness_Area4 = 1 THEN 4
        ELSE 0
   END AS wilderness_area

FROM public.covertype_test)
t1
GROUP BY 1, 3
ORDER BY 2 DESC
LIMIT 1;
```

La salida de la operación anterior tendrá un aspecto semejante al del siguiente ejemplo.

```
+----------------------+-------+-----------------+
| predicted_cover_type | count | wilderness_area |
+----------------------+-------+-----------------+
|                    2 | 15738 |               1 |
+----------------------+-------+-----------------+
```

### Temas relacionados de

Para obtener más información sobre Amazon Redshift ML, consulte la siguiente documentación:

- Costos de utilizar Amazon Redshift ML
- Operación CREATE MODEL
- Función EXPLAIN_MODEL

Para obtener más información sobre machine learning, consulte la siguiente documentación:

- Información general sobre machine learning
- Machine learning para principiantes y expertos
- ¿Qué es la equidad y la explicabilidad del modelo en las predicciones de machine learning?

## Integración de Amazon Redshift ML con Amazon Bedrock

En esta sección se describe cómo utilizar la integración de Amazon Redshift ML con Amazon Bedrock. Con esta característica, puede invocar un modelo de Amazon Bedrock mediante SQL y

puede utilizar los datos de un almacenamiento de datos de Amazon Redshift para crear aplicaciones de IA generativa, como la generación de texto, el análisis de opiniones o la traducción.

Temas


- Creación o actualización de un rol de IAM para la integración de Amazon Redshift ML con Amazon Bedrock
- Creación de un modelo externo para la integración de Amazon Redshift ML con Amazon Bedrock
- Uso de un modelo externo para la integración de Amazon Redshift ML con Amazon Bedrock
- Ingeniería de peticiones para la integración de Amazon Redshift ML con Amazon Bedrock


## Creación o actualización de un rol de IAM para la integración de Amazon Redshift ML con Amazon Bedrock

En esta sección se describe cómo crear un rol de IAM para usar con la integración de Amazon Redshift ML con Amazon Bedrock.

Agregue la siguiente política al rol de IAM que utilice con la integración de Amazon Redshift ML con Amazon Bedrock:

- `AmazonBedrockFullAccess`

Para permitir que Amazon Redshift asuma un rol para interactuar con otros servicios, agregue la siguiente política de confianza al rol de IAM:

```
{
   "Version": "2012-10-17",
   "Statement": [
      {
         "Effect": "Allow",
         "Principal": {
            "Service": [
               "redshift.amazonaws.com"
            ]
         },
         "Action": "sts:AssumeRole"
      }
   ]
}
```

Si el clúster o el espacio de nombres está en una VPC, siga las instrucciones que se indican en Agrupe y ajuste la configuración para la administración de Amazon Redshift ML.

Si necesita una política más restrictiva, puede crear una que incluya solo los permisos de Amazon Bedrock especificados en las siguientes páginas:

- Agrupe y ajuste la configuración para la administración de Amazon Redshift ML
- Permisos necesarios para usar machine learning (ML) de Amazon Redshift

Para obtener información sobre la creación de un rol de IAM, consulte Creación de un rol de IAM en la Guía del usuario de AWS Identity and Access Management.

## Creación de un modelo externo para la integración de Amazon Redshift ML con Amazon Bedrock

En esta sección se muestra cómo crear un modelo externo para usarlo como interfaz para Amazon Bedrock dentro del almacenamiento de datos de Amazon Redshift.

Para invocar un modelo de Amazon Bedrock desde Amazon Redshift, primero debe ejecutar el comando `CREATE EXTERNAL MODEL`. Este comando crea un objeto modelo externo en la base de datos y una función de usuario asociada que se utiliza para generar contenido de texto con Amazon Bedrock.

El siguiente ejemplo de código muestra un comando `CREATE EXTERNAL MODEL` básico:

```
CREATE EXTERNAL MODEL llm_claude
FUNCTION llm_claude_func
IAM_ROLE '<IAM role arn>'
MODEL_TYPE BEDROCK
SETTINGS (
   MODEL_ID 'anthropic.claude-v2:1',
   PROMPT 'Summarize the following text:');
```

El comando `CREATE EXTERNAL MODEL` tiene una interfaz unificada y coherente con Amazon Bedrock para todos los modelos fundacionales (FM) que admiten mensajes. Esta es la opción predeterminada cuando se usa el comando `CREATE EXTERNAL MODEL` o cuando se especifica explícitamente el tipo de solicitud que será `UNIFIED`. Para obtener más información, consulte la Documentación de la API de conversión en la Documentación de la API de Amazon Bedrock.

Si un FM no admite mensajes, debe establecer la configuración de `request_type` en `RAW`. Si establece `request_type` en `RAW`, debe crear la solicitud enviada a Amazon Bedrock cuando utilice la función de inferencia en función del FM seleccionado.

El parámetro `PROMPT` del comando `CREATE EXTERNAL MODEL` es una petición estática. Si necesita una petición dinámica para la aplicación, debe especificarla cuando utilice la función de inferencia. Para obtener más detalles, consulte Ingeniería de peticiones para la integración de Amazon Redshift ML con Amazon Bedrock siguiente.

Para obtener más información sobre la instrucción `CREATE EXTERNAL MODEL` y sus parámetros y configuración, consulte CREATE EXTERNAL MODEL.

## Uso de un modelo externo para la integración de Amazon Redshift ML con Amazon Bedrock

En esta sección se muestra cómo invocar un modelo externo para generar texto en respuesta a las peticiones proporcionadas. Para invocar un modelo externo, utilice la función de inferencia con la que ha creado `CREATE EXTERNAL MODEL`.

Temas

- Inferencia con modelos de tipo solicitud UNIFIED
- Inferencia con modelos de tipo solicitud RAW
- Funciones de inferencia como funciones específicas del principal
- Notas sobre el uso de la función de inferencia

### Inferencia con modelos de tipo solicitud **`UNIFIED`**

La función de inferencia para los modelos con tipo de solicitud `UNIFIED` tiene los tres parámetros siguientes que se transfieren a la función en orden:

- Texto de entrada (obligatorio): este parámetro especifica el texto de entrada que Amazon Redshift transfiere a Amazon Bedrock.
- Configuración de inferencia y campos de solicitud de modelo adicionales (opcionales): Amazon Redshift transfiere estos parámetros a los parámetros correspondientes de la API del modelo de conversión.

En el siguiente ejemplo de código se muestra cómo utilizar una función de inferencia de tipo `UNIFIED`:

```
SELECT llm_claude_func(input_text, object('temperature', 0.7, 'maxtokens', 500))
   FROM some_data;
```

### Inferencia con modelos de tipo solicitud **RAW**

La función de inferencia para los modelos con tipo de solicitud `RAW` solo tiene un parámetro de tipo de datos `SUPER`. La sintaxis de este parámetro depende del modelo de Amazon Bedrock utilizado.

En el siguiente ejemplo de código se muestra cómo utilizar una función de inferencia de tipo `RAW`:

```
SELECT llm_titan_func(
    object(
        "inputText", "Summarize the following text: " | input_text,
        "textGenerationConfig", object("temperature", 0.5, "maxTokenCount", 500)
    )
)
FROM some_data;
```

### Funciones de inferencia como funciones específicas del principal

Las funciones de inferencia de los modelos de Amazon Bedrock se pueden ejecutar como funciones exclusivas del nodo principal cuando la consulta que las utiliza no hace referencia a ninguna tabla. Esto puede resultar útil si quiere hacer una pregunta rápidamente a un LLM.

En el siguiente ejemplo de código se muestra cómo utilizar una función de inferencia específica del principal:

```
SELECT general_titan_llm_func('Summarize the benefits of LLM on data analytics in 100
 words');
```

### Notas sobre el uso de la función de inferencia

Tenga en cuenta lo siguiente cuando utilice funciones de inferencia con la integración de Amazon Redshift ML con Amazon Bedrock:

- Los nombres de los parámetros de todos los modelos de Amazon Bedrock distinguen entre mayúsculas y minúsculas. Si los parámetros no coinciden con los requeridos por el modelo, es posible que Amazon Bedrock los ignore discretamente.

- El rendimiento de las consultas de inferencia está limitado por las cuotas de tiempo de ejecución de los distintos modelos que ofrece Amazon Bedrock en las distintas regiones. Para obtener más información, consulte ¿Qué es Amazon Bedrock? en la Guía del usuario de Amazon Bedrock.
- Si necesita un rendimiento coherente y garantizado, considere la posibilidad de obtener un rendimiento aprovisionado para el modelo que necesita de Amazon Bedrock. Para obtener más información, consulte Aumentar la capacidad de invocación de modelos con el rendimiento aprovisionado en Amazon Bedrock en la Guía del usuario de Amazon Bedrock.
- Las consultas de inferencia con grandes cantidades de datos pueden tener excepciones de limitación. Esto se debe a las cuotas de tiempo de ejecución limitadas de Amazon Bedrock. Amazon Redshift vuelve a intentar las solicitudes varias veces, pero las consultas pueden seguir limitándose porque el rendimiento de los modelos no aprovisionados puede ser variable.
- Si encuentra excepciones que imponen limitaciones provenientes de Amazon Bedrock, por ejemplo, `Too many requests, please wait before trying again.`, incluso con pequeñas cantidades de datos, compruebe las cuotas en Service Quotas en su cuenta de Amazon Bedrock. Compruebe que la cuota de cuenta aplicada sea como mínimo igual al valor de la cuota predeterminada de AWS para las solicitudes de InvokeModel para el modelo que está utilizando.

## Ingeniería de peticiones para la integración de Amazon Redshift ML con Amazon Bedrock

En esta sección se muestra cómo utilizar las peticiones estáticas con un modelo externo.

Para utilizar peticiones de prefijo y sufijo estáticas para el modelo externo, indíquelas mediante los parámetros `PROMPT` y `SUFFIX` de la instrucción `CREATE EXTERNAL MODEL`. Estas peticiones se agregan a todas las consultas que utilizan el modelo externo.

En el siguiente ejemplo se muestra cómo agregar peticiones de prefijo y sufijo a un modelo externo:

```
CREATE EXTERNAL MODEL llm_claude
FUNCTION llm_claude_func
IAM_ROLE '<IAM role arn>'
MODEL_TYPE BEDROCK
SETTINGS (
   MODEL_ID 'anthropic.claude-v2:1',
   PROMPT 'Summarize the following text:',
   SUFFIX 'Respond in an analytic tone');
```

Para utilizar peticiones dinámicas, puede proporcionarlas al utilizar la función de inferencia concatenándolas en la entrada de la función. En el siguiente ejemplo se muestra cómo utilizar peticiones dinámicas con una función de inferencia:

```
SELECT llm_claude_func('Summarize the following review:' | input_text | 'The review
 should have formal tone.')
FROM some_data
```

# Ajuste del rendimiento de las consultas

Amazon Redshift utiliza consultas basadas en el lenguaje de consulta estructurada (SQL) para interactuar con los datos y los objetos en el sistema. El Data Manipulation Language (DML, Lenguaje de manipulación de datos) es el subconjunto de SQL que el usuario utiliza para ver, añadir, cambiar y eliminar datos. Data Definition Language (DDL, Lenguaje de definición de datos) es el subconjunto de SQL que el usuario utiliza para añadir, cambiar y eliminar objetos de la base de datos como tablas y vistas.

Una vez que su sistema está configurado, por lo general trabajará con DML la mayor parte del tiempo, en especial con el comando SELECT para recuperar y ver los datos. Para escribir consultas eficaces de recuperación de datos en Amazon Redshift, debe familiarizarse con el comando SELECT y aplicar los consejos descritos en Prácticas recomendadas de Amazon Redshift para el diseño de tablas para maximizar la eficiencia de las consultas.

Para comprender cómo Amazon Redshift procesa las consultas, utilice las secciones Procesamiento de consultas y Análisis y mejora de consultas. Luego, puede aplicar esta información junto con las herramientas de diagnóstico para identificar y eliminar problemas de rendimiento de las consultas.

Para identificar y abordar algunos de los problemas más frecuentes y graves que posiblemente enfrente con las consultas de Amazon Redshift, utilice la sección Solución de problemas de consultas.

Temas


- Procesamiento de consultas
- Análisis y mejora de consultas
- Solución de problemas de consultas


# Procesamiento de consultas

Amazon Redshift dirige una consulta SQL enviada a través del analizador y el optimizador para desarrollar un plan de consulta. El motor de ejecución luego traduce el plan de consulta en un código y lo envía a los nodos de computación para su ejecución.

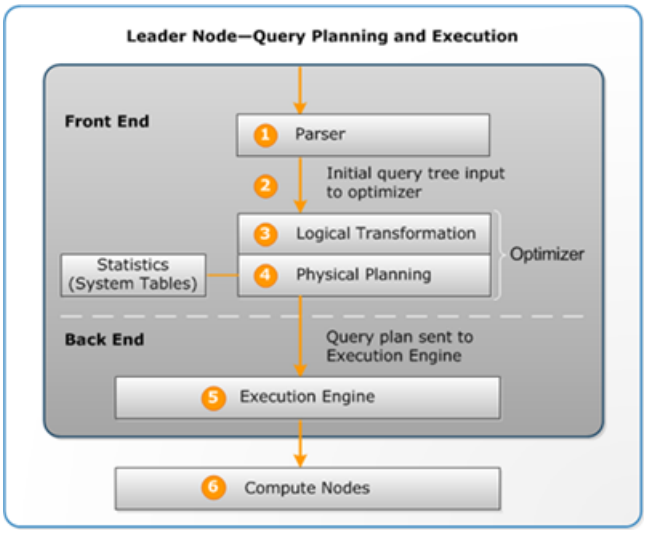
Temas


- Flujo de trabajo de planificación y ejecución de consultas
- Creación e interpretación de un plan de consultas

- Revisión de los pasos del plan de consulta
- Factores que afectan al rendimiento de las consultas

## Flujo de trabajo de planificación y ejecución de consultas

En el siguiente diagrama, se proporciona una vista general del flujo de trabajo de planificación y ejecución de consultas.




El flujo de trabajo de planificación y ejecución de consultas consta de los siguientes pasos:

1. El nodo principal recibe la consulta y analiza el SQL.
2. El analizador genera un árbol de consultas inicial que es una representación lógica de la consulta original. A continuación, Amazon Redshift inserta este árbol de consultas en el optimizador de consultas.
3. El optimizador evalúa y, de ser necesario, reescribe la consulta para maximizar su eficacia. Este proceso, a menudo, crea múltiples consultas relacionadas que reemplazan una misma consulta.
4. El optimizador genera un plan de consulta (o varios, si el paso anterior generó consultas múltiples) para ejecutarlo con el mejor rendimiento posible. El plan de consulta especifica las opciones de ejecución, como los tipos de combinaciones, el orden de las combinaciones, las opciones de agregación y los requisitos de distribución de datos.

   Puede utilizar el comando EXPLAIN para ver el plan de consulta. El plan de consulta es una herramienta fundamental para analizar y ajustar consultas complejas. Para obtener más información, consulte Creación e interpretación de un plan de consultas.
5. El motor de ejecución traduce el plan de consulta en pasos, segmentos y secuencias:

Paso

Cada paso es una operación individual que es necesario hacer durante la ejecución de consultas. Los pasos se pueden combinar para permitir que los nodos de computación realicen una consulta, una combinación y cualquier otra operación en la base de datos.

Segmento

Es una combinación de distintos pasos que se pueden realizar por un mismo proceso, además de ser la unidad mínima de compilación que puede ejecutar un sector de un nodo de computación. Un sector es la unidad de procesamiento en paralelo de Amazon Redshift. Los segmentos de una secuencia se ejecutan en paralelo.

De streaming

Es un conjunto de segmentos que se reparten entre los sectores disponibles de un nodo de computación.

El motor de ejecución genera un código compilado en función de los pasos, los segmentos y las secuencias. El código compilado se ejecuta más rápido que el código interpretado y utiliza menor capacidad informática. Luego, este código compilado se difunde a los nodos de computación.

> **Note**
>
> Al realizar análisis comparativos de sus consultas, siempre debe comparar los tiempos de la segunda ejecución de una consulta, porque el tiempo de la primera ejecución incluye el trabajo adicional de compilar el código. Para obtener más información, consulte Factores que afectan al rendimiento de las consultas.

6. Los sectores del nodo informático ejecutan los segmentos de la consulta en paralelo. Como parte de este proceso, Amazon Redshift aprovecha la comunicación optimizada de red, la memoria y la administración de discos para transmitir los resultados intermedios de un paso del plan de consulta al siguiente. Esto también ayuda a agilizar la ejecución de consultas.

Los pasos 5 y 6 tienen lugar una sola vez en cada secuencia. El motor crea segmentos ejecutables para una secuencia y los envía a los nodos de computación. Cuando los segmentos de esa secuencia están completos, el motor genera los segmentos para la próxima secuencia. De esta manera, el motor puede analizar lo que ocurrió la secuencia anterior (por ejemplo, si las operaciones estaban basadas en el disco) para influir sobre la generación de segmentos en la próxima secuencia.

Cuando los nodos de computación están listos, devuelven los resultados de las consultas al nodo principal para el procesamiento final. El nodo principal fusiona los datos en un conjunto único de resultados y realiza las tareas de ordenación o agregación que sean necesarias. Luego, el nodo principal devuelve los resultados al cliente.

> **Note**
>
> Es probable que los nodos de computación devuelvan algunos datos al nodo principal durante la ejecución de consultas si fuera necesario. Por ejemplo, si tiene una subconsulta con una cláusula LIMIT, el límite se aplica al nodo principal antes de que se distribuyan los datos en todo el clúster para seguir trabajando con ellos.

## Creación e interpretación de un plan de consultas

Puede utilizar el plan de consulta para obtener información sobre las operaciones individuales necesarias para ejecutar una consulta. Antes de trabajar con un plan de consulta, le recomendamos que primero comprenda cómo Amazon Redshift administra el procesamiento de las consultas y la creación de planes de consultas. Para obtener más información, consulte Flujo de trabajo de planificación y ejecución de consultas.

Para crear un plan de consulta, ejecute el comando EXPLAIN seguido del texto real de la consulta. En el plan de consulta, se proporciona la siguiente información:

- Las operaciones que realizará el motor de ejecución, leyendo los resultados de abajo arriba.
- El tipo de paso que realiza cada operación.
- Las tablas y las columnas que se utilizan en cada operación.
- La cantidad de datos que se procesa en cada operación, en cuanto a la cantidad de filas y al ancho de datos en bytes.
- El costo relativo de la operación. El costo es una medida que compara los tiempos relativos de ejecución de los pasos de un plan. El costo no proporciona información precisa acerca de los tiempo de ejecución reales ni sobre el consumo de memoria, como tampoco proporciona una comparación significativa entre los planes de ejecución. El costo le da una idea de cuáles son las operaciones de una consulta que están consumiendo la mayor cantidad de recursos.

El comando EXPLAIN no ejecuta propiamente la consulta. Solo muestra el plan que ejecuta Amazon Redshift si la consulta se lleva a cabo en las condiciones operativas actuales. Si cambia el

esquema o los datos de una tabla y ejecuta nuevamente el comando ANALYZE para actualizar los metadatos estadísticos, el plan de consulta podría ser diferente.

La salida del plan de consulta que genera el comando EXPLAIN es una vista general simplificada de la ejecución de consultas. No muestra los detalles del procesamiento en paralelo de las consultas. Para ver información detallada, ejecute la consulta en cuestión y, luego, obtenga información resumida de la consulta desde la vista SVL_QUERY_SUMMARY o SVL_QUERY_REPORT. Para obtener más información acerca de cómo usar estas vistas, consulte Análisis del resumen de consultas.

En el siguiente ejemplo, se muestra una salida del comando EXPLAIN para una consulta simple GROUP BY sobre la tabla EVENT:

```
explain select eventname, count(*) from event group by eventname;

                            QUERY PLAN
------------------------------------------------------------------
XN HashAggregate  (cost=131.97..133.41 rows=576 width=17)
  ->  XN Seq Scan on event  (cost=0.00..87.98 rows=8798 width=17)
```

EXPLAIN devuelve las siguientes métricas para cada operación:

Costo

Un valor relativo que es útil para comparar operaciones dentro de un plan. El costo se compone de dos valores decimales separados por dos puntos, por ejemplo `cost=131.97..133.41`. El primer valor, en este caso 131.97, proporciona el costo relativo del retorno de la primera fila de esta operación. El segundo valor, en este caso 133.41, proporciona el costo relativo de completar la operación. Los costos del plan de consulta se van acumulando a medida que estudia el plan, por lo que el costo HashAggregate de este ejemplo (131.97..133.41) incluye el costo del examen secuencial que está debajo (0.00..87.98).

Filas

La cantidad estimada de filas que se devuelven. En este ejemplo, se espera que el examen devuelva 8798 filas. Se espera que el operador HashAggregate por sí solo devuelva 576 filas (después de descartar los nombres de eventos duplicados del conjunto de resultados).

> **Note**
>
> El cálculo de filas se basa en las estadísticas disponibles que genera el comando ANALYZE. Si ANALYZE no se ejecutó recientemente, el cálculo será menos fiable.

Ancho

El ancho estimado de la fila promedio, medido en bytes. En este ejemplo, se espera que la fila promedio sea de 17 bytes de ancho.

## Operadores EXPLAIN

En esta sección, se describen brevemente los operadores que ven con mayor frecuencia en la salida del comando EXPLAIN. Para obtener una lista completa de los operadores, consulte EXPLAIN en la sección sobre comandos SQL.

Operador de examen secuencial

El operador de examen secuencial (Seq Scan) indica el examen de una tabla. Seq Scan examina cada columna de la tabla de manera secuencial, de principio a fin, y evalúa las restricciones de la consulta (en la cláusula WHERE) para cada una de las filas.

Operadores de combinación

Amazon Redshift selecciona los operadores de combinación en función del diseño físico de las tablas que se combinarán, la ubicación de los datos que se necesitan para la combinación y los requisitos específicos de la propia consulta.

- Bucle anidado

  La combinación menos óptima es el bucle anidado, que se utiliza principalmente para las combinaciones cruzadas (productos cartesianos) y para algunas combinaciones de desigualdades.
- Operadores hash join y hash

  Se utilizan los operadores hash join y hash, por lo general más rápidos que la combinación de bucle anidado, para las combinaciones internas y las combinaciones externas izquierda y derecha. Estos operadores se utilizan para combinar tablas en las que las columnas de combinación no son claves de distribución ni claves de ordenación al mismo tiempo. El operador hash crea una tabla

hash para la tabla interna de la combinación; el operador hash join lee la tabla exterior, aplica la función hash a la columna de combinación y encuentra coincidencias en la tabla hash interna.

- Merge Join

  Se utiliza el operador merge join, que, por lo general, es el operador más rápido, para las combinaciones internas y externas. Merge join no se utiliza para las combinaciones totales. Este operador se utiliza para combinar tablas en las que las columnas de combinación son claves de distribución y claves de ordenación al mismo tiempo, y cuando menos del 20 por ciento de las tablas de combinación está desordenado. Lee dos tablas ordenadas en orden y encuentra las filas que coinciden. Para ver el porcentaje de filas desordenadas, consulta la tabla de sistema SVV_TABLE_INFO.

- Unión espacial

  Suele ser una unión rápida basada en la proximidad de los datos espaciales, que se utiliza para los tipos de datos `GEOMETRY` y `GEOGRAPHY`.

### Operadores de agregación

El plan de consulta utiliza los siguientes operadores en las consultas que implican funciones de agregación y operaciones GROUP BY.

- Aggregate

  Operador para las funciones de agregación escalar, como AVG y SUM.

- HashAggregate

  Operador para las funciones de agregación agrupadas desordenadas.

- GroupAggregate

  Operador para las funciones de agregación agrupadas ordenadas.

### Operadores de ordenación

El plan de consulta utiliza los siguientes operadores cuando las consultas tienen que ordenar o fusionar conjuntos de resultados.

- Sort

Evalúa la cláusula ORDER BY y otras operaciones de ordenación, como las que se necesitan en las consultas y combinaciones UNION, en las consultas SELECT DISTINCT y en las funciones de ventana.

- Merge

  Produce resultados ordenados finales conforme a los resultados intermedios ordenados que se derivan de las operaciones en paralelo.

### Operadores UNION, INTERSECT y EXCEPT

El plan de consulta utiliza los siguientes operadores en las consultas que implican operaciones de conjunto con los operadores UNION, INTERSECT y EXCEPT.

- Operador subquery

  Se utiliza para ejecutar consultas UNION.
- Hash Intersect Distinct

  Se utiliza para ejecutar consultas INTERSECT .
- Operador SetOp Except

  Se utiliza para ejecutar consultas EXCEPT (o MINUS).

### Otros operadores

Los siguientes operadores también aparecen con frecuencia en la salida EXPLAIN para las consultas rutinarias.

- Unique

  Elimina los duplicados de las consultas SELECT DISTINCT y UNION.
- Límite

  Procesa la cláusula LIMIT.
- Window

  Ejecuta las funciones de ventana.
- Resultado

Ejecuta las funciones escalares que no implican obtener acceso a ninguna tabla.

- Subplan

  Se utiliza para algunas subconsultas.

- Red

  Envía los resultados intermedios al nodo principal para seguir trabajando con ellos.

- Materialize

  Guarda las filas para entradas de combinaciones de bucle anidado y en algunas de combinaciones de fusión.

## Combinaciones con EXPLAIN

El optimizador de consultas utiliza diferentes tipos de combinaciones para recuperar datos de tablas, en función de la estructura de la consulta y de las tablas subyacentes. La salida EXPLAIN indica el tipo de combinación, las tablas utilizadas y la manera en la que se distribuyen los datos de la tabla en todo el clúster para describir cómo se procesa la consulta.

Ejemplos de tipos de combinación

En los siguientes ejemplos, se muestran los distintos tipos de combinación que puede utilizar el optimizador de consultas. El tipo de combinación que se utiliza en el plan de consulta depende del diseño físico de las tablas implicadas.

Ejemplo: Combinación de dos tablas mediante el operador hash

La siguiente consulta combina EVENT y CATEGORY en la columna CATID. CATID es la clave de distribución y ordenación para CATEGORY, pero no para EVENT. Se realiza una combinación hash con EVENT como la tabla externa y CATEGORY como la tabla interna. Como CATEGORY es la tabla más pequeña, el planificador difunde una copia de esta en los nodos de computación durante el procesamiento de la consulta mediante el comando DS_BCAST_INNER. El costo de combinación en este ejemplo contempla la mayor parte del costo acumulado del plan.

```
explain select * from category, event where category.catid=event.catid;

                               QUERY PLAN
-----------------------------------------------------------------------
```

```
XN Hash Join DS_BCAST_INNER  (cost=0.14..6600286.07 rows=8798 width=84)
  Hash Cond: ("outer".catid = "inner".catid)
  ->  XN Seq Scan on event  (cost=0.00..87.98 rows=8798 width=35)
  ->  XN Hash  (cost=0.11..0.11 rows=11 width=49)
        ->  XN Seq Scan on category  (cost=0.00..0.11 rows=11 width=49)
```

> **Note**
>
> Las sangrías alineadas de los operadores en la salida EXPLAIN, a menudo, indican que esas operaciones no dependen unas de otras y que pueden comenzar en paralelo. En el ejemplo anterior, si bien el examen de la tabla EVENT y la operación hash están alineados, el examen EVENT debe esperar hasta que la operación hash haya terminado por completo.

Ejemplo: Combinación de dos tablas mediante el operador merge

La siguiente consulta también usa el operador SELECT*, pero combina SALES y LISTING en la columna LISTID, donde LISTID está configurada como la clave de distribución y de ordenación para ambas tablas. Se elige una combinación de fusión y no se necesita redistribuir datos para la combinación (DS_DIST_NONE).

```
explain select * from sales, listing where sales.listid = listing.listid;
QUERY PLAN
------------------------------------------------------------------------------
XN Merge Join DS_DIST_NONE  (cost=0.00..6285.93 rows=172456 width=97)
  Merge Cond: ("outer".listid = "inner".listid)
  ->  XN Seq Scan on listing  (cost=0.00..1924.97 rows=192497 width=44)
  ->  XN Seq Scan on sales  (cost=0.00..1724.56 rows=172456 width=53)
```

En el siguiente ejemplo, se demuestran los distintos tipos de combinación dentro de la misma consulta. Como en el ejemplo anterior, SALES y LISTING están combinadas por fusión, pero la tercera tabla, EVENT, debe estar combinada por hash con los resultados de la combinación de fusión. Una vez más, la combinación hash implica un costo de difusión.

```
explain select * from sales, listing, event
where sales.listid = listing.listid and sales.eventid = event.eventid;
                                  QUERY PLAN
------------------------------------------------------------------------------
XN Hash Join DS_BCAST_INNER  (cost=109.98..3871130276.17 rows=172456 width=132)
  Hash Cond: ("outer".eventid = "inner".eventid)
```

```
  ->  XN Merge Join DS_DIST_NONE  (cost=0.00..6285.93 rows=172456 width=97)
        Merge Cond: ("outer".listid = "inner".listid)
        ->  XN Seq Scan on listing  (cost=0.00..1924.97 rows=192497 width=44)
        ->  XN Seq Scan on sales  (cost=0.00..1724.56 rows=172456 width=53)
  ->  XN Hash  (cost=87.98..87.98 rows=8798 width=35)
        ->  XN Seq Scan on event  (cost=0.00..87.98 rows=8798 width=35)
```

Ejemplo: Operadores join, aggregate y sort

La siguiente consulta ejecuta una unión hash de las tablas SALES y EVENT, seguida de las operaciones de agregación y ordenación para justificar la función agrupada SUM y la cláusula ORDER BY. El operador inicial sort se ejecuta en paralelo en los nodos de computación. Luego, el operador network envía los resultados al nodo principal, donde el operador merge produce resultados finales ordenados.

```
explain select eventname, sum(pricepaid) from sales, event
where sales.eventid=event.eventid group by eventname
order by 2 desc;
                                          QUERY PLAN
------------------------------------------------------------------------------
 XN Merge  (cost=1002815366604.92..1002815366606.36 rows=576 width=27)
  Merge Key: sum(sales.pricepaid)
  ->  XN Network  (cost=1002815366604.92..1002815366606.36 rows=576 width=27)
        Send to leader
        ->  XN Sort  (cost=1002815366604.92..1002815366606.36 rows=576 width=27)
              Sort Key: sum(sales.pricepaid)
              ->  XN HashAggregate  (cost=2815366577.07..2815366578.51 rows=576
 width=27)
                    ->  XN Hash Join DS_BCAST_INNER  (cost=109.98..2815365714.80
 rows=172456 width=27)
                          Hash Cond: ("outer".eventid = "inner".eventid)
                          ->  XN Seq Scan on sales  (cost=0.00..1724.56 rows=172456
 width=14)
                          ->  XN Hash  (cost=87.98..87.98 rows=8798 width=21)
                                ->  XN Seq Scan on event  (cost=0.00..87.98 rows=8798
 width=21)
```

Redistribución de datos

La salida EXPLAIN para las combinaciones también especifica un método con el cual los datos se mueven en torno a un clúster para facilitar la combinación. Este movimiento de datos puede ser una difusión o bien una redistribución. En una difusión, los valores de los datos de un lado de

la combinación se copian desde cada nodo de computación a los demás nodos de computación de manera que cada nodo de computación tenga una copia completa de los datos. En una redistribución, se envían los valores de los datos participantes desde el sector donde están hacia un nuevo sector (posiblemente en un nodo diferente). Los datos, por lo general, se redistribuyen para coincidir con la clave de distribución de la otra tabla participante de la combinación si esa clave de distribución es una de las columnas que se está combinando. Si ninguna de las tablas tiene claves de distribución en las columnas que se están combinando, se distribuyen ambas tablas, o bien se difunde la tabla interna en cada uno de los nodos.

La salida EXPLAIN también hace referencia a las tablas internas y externas. Primero se examina la tabla interna; esta aparece cerca de la parte inferior del plan de consulta. La tabla interna es la tabla en la que se buscan coincidencias. Por lo general, se almacena en la memora y suele ser la tabla origen de algoritmos hash y, de ser posible, la tabla más pequeña de las dos que se están combinando. La tabla externa es el origen de las filas en las que se busca coincidencias con la tabla interna. Por lo general, se lee del disco. El optimizador de consultas selecciona la tabla interna y externa en función de las estadísticas de la base de datos extraídas de la última ejecución del comando ANALYZE. El orden de las tablas en la cláusula FROM de una consulta no determina cuál es la tabla interna y cuál es la externa.

Utilice los siguientes atributos en los planes de consulta para identificar cómo se mueven los datos y, así, facilitar una consulta:

- DS_BCAST_INNER

  Se difunde una copia de toda la tabla interna a todos los nodos de computación.
- DS_DIST_ALL_NONE

  No es necesario redistribuir, porque ya se distribuyó la tabla interna a cada nodo con el atributo DISTSTYLE ALL.
- DS_DIST_NONE

  No se redistribuye ninguna tabla. Es posible que haya combinaciones colocadas porque se combinan los sectores correspondientes sin mover los datos entre los nodos.
- DS_DIST_INNER

  Se redistribuye la tabla interna.
- DS_DIST_OUTER

  Se redistribuye la tabla externa.

- DS_DIST_ALL_INNER

  Se redistribuye toda la tabla interna a un único sector porque la tabla externa utiliza el atributo DISTSTYLE ALL.
- DS_DIST_BOTH

  Ambas tablas se redistribuyen.

## Revisión de los pasos del plan de consulta

Para ver los pasos de un plan de consulta, ejecute el comando EXPLAIN. En el siguiente ejemplo, se muestra una consulta SQL y se explica el resultado. Si lee el plan de consulta desde abajo hacia arriba, puede ver cada una de las operaciones lógicas que se utilizan para realizar la consulta. Para obtener más información, consulte Creación e interpretación de un plan de consultas.

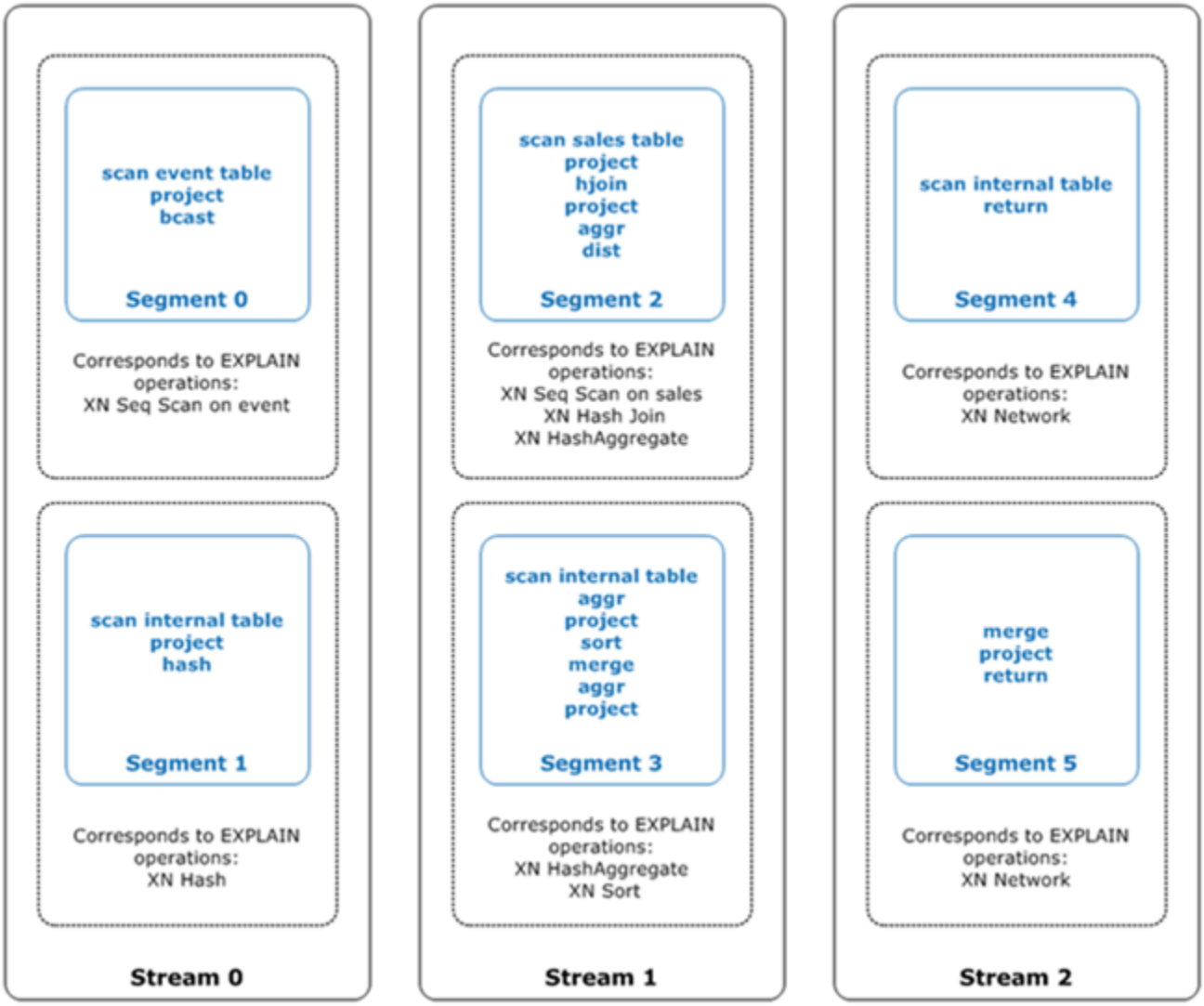
```
explain
select eventname, sum(pricepaid) from sales, event
where sales.eventid = event.eventid
group by eventname
order by 2 desc;
```

```
XN Merge  (cost=1002815366604.92..1002815366606.36 rows=576 width=27)
  Merge Key: sum(sales.pricepaid)
  ->  XN Network  (cost=1002815366604.92..1002815366606.36 rows=576 width=27)
        Send to leader
        ->  XN Sort  (cost=1002815366604.92..1002815366606.36 rows=576 width=27)
              Sort Key: sum(sales.pricepaid)
              ->  XN HashAggregate  (cost=2815366577.07..2815366578.51 rows=576
 width=27)
                    ->  XN Hash Join DS_BCAST_INNER  (cost=109.98..2815365714.80
 rows=172456 width=27)
                          Hash Cond: ("outer".eventid = "inner".eventid)
                          ->  XN Seq Scan on sales  (cost=0.00..1724.56 rows=172456
 width=14)
                          ->  XN Hash  (cost=87.98..87.98 rows=8798 width=21)
                                ->  XN Seq Scan on event  (cost=0.00..87.98 rows=8798
 width=21)
```

Como parte de la generación de un plan de consulta, el optimizador de consultas desglosa el plan en flujos, segmentos y pasos. El optimizador de consultas desglosa el plan para prepararse para

distribuir los datos y la carga de trabajo de consulta a los nodos informáticos. Para obtener más información acerca de los flujos, segmentos y pasos, consulte Flujo de trabajo de planificación y ejecución de consultas.

En la siguiente ilustración se muestra la consulta anterior y el plan de consulta asociado. Muestra cómo las operaciones de consulta involucradas se asignan a los pasos que Amazon Redshift utiliza para generar código compilado para los sectores de nodos informáticos. Cada operación del plan de consulta se asigna a distintos pasos dentro de los segmentos y, a menudo, a distintos segmentos dentro de las secuencias.




En esta ilustración, el optimizador de consultas ejecuta el plan de consultas de la siguiente manera:

1. En `Stream 0`, la consulta ejecuta `Segment 0` con una operación de análisis secuencial para analizar la tabla `events`. La consulta continúa `Segment 1` con una operación hash para crear la tabla hash para la tabla interna de la combinación.

2. En `Stream 1`, la consulta ejecuta `Segment 2` con una operación de análisis secuencial para analizar la tabla `sales`. Continúa `Segment 2` con una combinación hash para unir tablas donde las columnas de unión no son claves de distribución y claves de ordenación a la vez. De nuevo continúa con `Segment 2` con un hash de agregación para agregar resultados. A continuación, la consulta ejecuta `Segment 3` con una operación de agregado hash para realizar funciones de agregación agrupadas no ordenadas y una operación de ordenación para evaluar la cláusula ORDER BY y otras operaciones de ordenación.
3. En `Stream 2`, la consulta ejecuta una operación de red en `Segment 4` y `Segment 5` para enviar resultados intermedios al nodo de directriz para su posterior procesamiento.

El último segmento de una consulta devuelve los datos. Si el conjunto de devoluciones se agrega u ordena, los nodos de cálculo envían cada uno su parte del resultado intermedio al nodo de directriz. A continuación, el nodo líder combina los datos para que el resultado final se pueda enviar de nuevo al cliente solicitante.

Para obtener más información acerca de los operadores EXPLAIN, consulte EXPLAIN.

## Factores que afectan al rendimiento de las consultas

Hay una serie de factores que pueden afectar al rendimiento de las consultas. Los siguientes aspectos de sus datos, su clúster y de las operaciones de la base de datos influyen en el tiempo de procesamiento de sus consultas.

- Cantidad de nodos, procesadores o sectores: los nodos informáticos están divididos en sectores. Si hay más nodos, hay más procesadores y más sectores lo que permite procesar sus consultas con mayor rapidez al ejecutar porciones de una consulta de manera simultánea en todos los sectores. Sin embargo, si hay más nodos también hay más gastos, por lo que necesita encontrar un equilibrio entre el costo y el rendimiento que sea adecuado para su sistema. Para obtener más información acerca de la arquitectura de clústeres de Amazon Redshift, consulte Arquitectura del sistema de almacenamiento de datos.
- Tipos de nodos: un clúster de Amazon Redshift puede utilizar uno de los varios tipos de nodos. Cada tipo de nodo ofrece diferentes tamaños y límites para ayudarlo a escalar su clúster de manera adecuada. El tamaño del nodo determina la capacidad de almacenamiento, la memoria, la CPU y el precio de cada nodo del clúster. Para obtener más información sobre los tipos de nodos, consulte Información general de los clústeres de Amazon Redshift en la Guía de administración de Amazon Redshift.

- Distribución de datos: Amazon Redshift almacena los datos de las tablas en los nodos informáticos en función del estilo de distribución de la tabla. Cuando ejecuta una consulta, el optimizador de consultas redistribuye los datos a los nodos informáticos según se necesite para realizar uniones y agregaciones. Elegir el estilo de distribución correcto para una tabla ayuda a reducir el impacto del paso de redistribución al localizar los datos en el lugar que deben estar, antes de que se realicen las combinaciones. Para obtener más información, consulte Distribución de datos para la optimización de consultas.
- Ordenación de los datos: Amazon Redshift almacena los datos de las tablas en el disco en un determinado orden, en función de las claves de ordenación de una tabla. El optimizador y el procesador de consultas utilizan la información de ubicación de los datos para reducir la cantidad de bloques que se deben examinar y, por ende, mejorar la velocidad de la consulta. Para obtener más información, consulte Claves de clasificación.
- Tamaño del conjunto de datos: un mayor volumen de datos en el clúster puede ralentizar el rendimiento de las consultas, ya que es necesario examinar y redistribuir más filas. Puede mitigar este efecto si limpia y archiva periódicamente los datos y si utiliza un predicado para restringir el conjunto de datos de las consultas.
- Operaciones simultáneas: ejecutar distintas operaciones a la vez puede afectar el rendimiento de las consultas. Cada operación utiliza uno o más slots de una cola de consultas disponible y utiliza la memora asociada a esos slots. Si hay otras operaciones en ejecución, es posible que no haya suficientes slots de colas de consulta disponibles. En este caso, la consulta tiene que esperar hasta que se abren slots para empezar a procesarse. Para obtener más información acerca de la creación y la configuración de colas de consultas, consulte Administración de la carga de trabajo.
- Estructura de la consulta: la forma en la que está escrita la consulta afecta el rendimiento. Siempre que sea posible, escriba las consultas para que procesen y devuelvan la menor cantidad de datos que satisfagan sus necesidades. Para obtener más información, consulte Prácticas recomendadas de Amazon Redshift para el diseño de consultas.
- Compilación de código: Amazon Redshift genera y compila código para cada plan de ejecución de consultas.

  El código compilado se ejecuta más rápido porque elimina la sobrecarga de tener que recurrir a un intérprete. Siempre tendrá algún costo adicional la primera vez que genere y compile código. Por este motivo, el rendimiento de la primera vez que ejecuta una consulta puede ser confuso. El costo adicional puede ser particularmente evidente cuando ejecuta consultas aisladas. Ejecute la consulta por segunda vez para determinar su rendimiento habitual. Amazon Redshift utiliza un servicio de compilación sin servidor para escalar las compilaciones de consultas más allá de los recursos informáticos de un clúster de Amazon Redshift. Los segmentos de código compilado

se almacenan en caché de manera local en el clúster y en una caché virtualmente ilimitada. Esta caché se conserva después del reinicio del clúster. Las ejecuciones posteriores de la misma consulta se realizan en menos tiempo, ya que se puede omitir la fase de compilación.

La caché no es compatible con todas las versiones de Amazon Redshift, por lo que la caché de compilación se vacía y el código se vuelve a compilar cuando se ejecutan consultas después de una actualización de la versión. Si las consultas tienen acuerdos de nivel de servicio (SLA) estrictos, le recomendamos que ejecute previamente segmentos de consulta que escaneen los datos de las tablas de clústeres. Esto permite que Amazon Redshift almacene en caché los datos de la tabla base, lo que reduce el tiempo de planificación de las consultas tras la actualización de una versión. Mediante el uso de un servicio de compilación escalable, Amazon Redshift puede compilar código en paralelo para proporcionar un rendimiento constante y rápido. La magnitud de la aceleración de la carga de trabajo depende de la complejidad y la simultaneidad de las consultas.

# Análisis y mejora de consultas

Recuperar información de un almacenamiento de datos de Amazon Redshift implica ejecutar consultas complejas en cantidades muy grandes de datos, cuyo procesamiento puede tomar mucho tiempo. Para garantizar que las consultas se procesen lo más rápido posible, hay una serie de herramientas que puede utilizar para identificar posibles problemas de rendimiento.

Temas


- Flujo de trabajo de análisis de consultas
- Revisión de alertas de consultas
- Análisis del plan de consulta
- Análisis del resumen de consultas
- Mejora del rendimiento de las consultas
- Consultas de diagnóstico para el ajuste de consultas


## Flujo de trabajo de análisis de consultas

Si una consulta está tardando más de lo previsto, utilice los siguientes pasos para identificar y corregir los problemas que podrían estar afectando negativamente el rendimiento de la consulta. Si no está seguro de qué consultas de su sistema podrían beneficiarse al ajustar su rendimiento,

comience por ejecutar una consulta de diagnóstico en Identificación de consultas que deben ajustarse con prioridad.

1. Asegúrese de que sus tablas estén diseñadas conforme a las prácticas recomendadas. Para obtener más información, consulte Prácticas recomendadas de Amazon Redshift para el diseño de tablas.
2. Considere eliminar o archivar cualquier dato innecesario que haya en sus tablas. Por ejemplo, suponga que sus consultas siempre incluyen los datos de los últimos 6 meses, pero tiene los datos de los últimos 18 meses en sus tablas. En este caso, puede eliminar o archivar los datos más antiguos para reducir la cantidad de registros que se deben examinar y distribuir.
3. Ejecute el comando VACUUM en las tablas de la consulta para recuperar espacio y reordene las filas. Ejecutar el comando VACUUM es útil si la región desordenada es grande y si la consulta usa la clave de ordenación en una combinación o en el predicado.
4. Ejecute el comando ANALYZE en las tablas de la consulta para asegurarse de que las estadísticas están actualizadas. Ejecutar el comando ANALYZE es útil si alguna de las tablas de la consulta cambió considerablemente su tamaño en el último tiempo. Si la ejecución de un comando ANALYZE completo tardada mucho tiempo, ejecútelo en una única columna para reducir el tiempo de procesamiento. Este método sigue actualizando las estadísticas de tamaño de la tabla, que es un factor importante en la planificación de consultas.
5. Asegúrese de haber ejecutado su consulta una vez por cada tipo de cliente (en función del tipo de protocolo de conexión que utiliza el cliente) para que la consulta se compile y almacene en caché. Este enfoque acelera las ejecuciones posteriores de la consulta. Para obtener más información, consulte Factores que afectan al rendimiento de las consultas.
6. Revise la tabla STL_ALERT_EVENT_LOG para identificar y corregir posibles problemas con su consulta. Para obtener más información, consulte Revisión de alertas de consultas.
7. Ejecute el comando EXPLAIN para obtener el plan de consulta y utilícelo para optimizar la consulta. Para obtener más información, consulte Análisis del plan de consulta.
8. Utilice las vistas SVL_QUERY_SUMMARY y SVL_QUERY_REPORT para obtener información resumida y utilícela para optimizar la consulta. Para obtener más información, consulte Análisis del resumen de consultas.

A veces, una consulta que se debe ejecutar rápidamente se ve obligada a esperar a que termine otra consulta que demanda mayor tiempo de ejecución. En ese caso, quizás no hay nada de la propia consulta que pueda mejorar, pero sí puede mejorar el rendimiento general del sistema al crear y utilizar colas de consultas para diferentes tipos de consultas. Para que se haga una idea de los

tiempos de espera de las consultas, consulte Revisión de los tiempos de espera de las colas de consultas. Para obtener más información acerca de la configuración de colas de consultas, consulte Administración de la carga de trabajo.

## Revisión de alertas de consultas

Para utilizar la tabla de sistema STL_ALERT_EVENT_LOG a fin de identificar y corregir posibles problemas de rendimiento con su consulta, siga los pasos a continuación:

1. Ejecute la siguiente función para determinar el ID de su consulta:

```
select query, elapsed, substring
from svl_qlog
order by query
desc limit 5;
```

   Examine el texto truncado de la consulta en el campo `substring` para determinar qué valor de `query` seleccionar. Si ejecutó la consulta más de una vez, utilice el valor `query` de la fila con el valor de `elapsed` más bajo. Esa es la fila de la versión compilada. Si ha ejecutado distintas consultas, puede aumentar el valor que utilizó la cláusula LIMIT para asegurarse de que su consulta esté incluida.

2. Seleccione las filas de STL_ALERT_EVENT_LOG para su consulta:

```
Select * from stl_alert_event_log where query = MyQueryID;
```

| userid | query | slice | segment | step | pid | xid | event | solution | event_time |
|---|---|---|---|---|---|---|---|---|---|
| 100 | 32359 | 4 | 0 | 0 | 8780 | 71195 | Very selective query filter:ratio=rows(2)/r | Review the choice of sort key to enable... | 2015-02-10 17:40:50 |
| 100 | 32359 | 5 | 0 | 0 | 8781 | 71195 | Very selective query filter:ratio=rows(2)/r | Review the choice of sort key to enable... | 2015-02-10 17:40:50 |
| 100 | 109142 | 4 | 0 | 0 | 8780 | 302411 | Very selective query filter:ratio=rows(2)/r | Review the choice of sort key to enable... | 2015-02-24 20:32:28 |
| 100 | 109142 | 5 | 0 | 0 | 8781 | 302411 | Very selective query filter:ratio=rows(2)/r | Review the choice of sort key to enable... | 2015-02-24 20:32:28 |
| 100 | 109828 | 4 | 1 | 0 | 8746 | 304543 | Very selective query filter:ratio=rows(3)/r | Review the choice of sort key to enable... | 2015-02-24 23:27:52 |
| 100 | 109828 | 5 | 1 | 0 | 8747 | 304543 | Very selective query filter:ratio=rows(3)/r | Review the choice of sort key to enable... | 2015-02-24 23:27:52 |
| 100 | 109829 | 4 | 1 | 0 | 8760 | 304543 | Very selective query filter:ratio=rows(3)/r | Review the choice of sort key to enable... | 2015-02-24 23:28:01 |
| 100 | 109829 | 5 | 1 | 0 | 8761 | 304543 | Very selective query filter:ratio=rows(3)/r | Review the choice of sort key to enable... | 2015-02-24 23:28:01 |
| 100 | 113910 | 4 | 1 | 0 | 8774 | 316848 | Very selective query filter:ratio=rows(3)/r | Review the choice of sort key to enable... | 2015-02-25 17:14:58 |

3. Evalúe los resultados de su consulta. Utilice la siguiente tabla para identificar posibles soluciones para cualquiera de los problemas que haya identificado.

> **Note**
>
> No todas las consultas tienen filas en STL_ALERT_EVENT_LOG, solo en las que se hayan identificado problemas.

| Problema | Valor del evento | Valor de la solución | Solución recomendada |
| --- | --- | --- | --- |
| Faltan estadísticas sobre las tablas de la consulta o están desactualizadas. | Faltan estadísticas para el planificador de consultas. | Ejecute el comando ANALYZE. | Consulte Faltan estadísticas de tablas o están desactualizadas. |
| Hay una combinación de bucle anidado (la combinación menos óptima) en el plan de consulta. | Hay una combinación de bucle anidado en el plan de consulta. | Revise los predicados de la combinación para evitar productos cartesianos. | Consulte Bucle anidado. |
| El examen omitió una cantidad considerablemente grande de filas que están marcadas como eliminadas pero no limpiadas o filas que fueron insertadas pero no confirmadas. | Se examinó una gran cantidad de filas eliminadas. | Ejecute el comando VACUUM para reclamar espacio eliminado. | Consulte Filas fantasma o filas sin confirmar. |
| Más de 1 000 000 filas fueron redistribuidas para una combinación hash o una agregación. | Se distribuyó una gran cantidad de filas en toda la red: las filas RowCount se distribuyeron para poder procesar la agregación. | Revise la elección de la clave de distribución para colocar la combinación o la agregación. | Consulte Distribución de datos poco óptima. |

| Problema | Valor del evento | Valor de la solución | Solución recomendada |
|---|---|---|---|
| Más de 1 000 000 filas fueron difundidas para una combinación hash. | Se difundió una gran cantidad de filas en toda la red. | Revise la elección de la clave de distribución para colocar la combinación y considere la opción de usar tablas distribuidas. | Consulte Distribución de datos poco óptima. |
| En el plan de consulta se indicó un estilo de redistribución DS_DIST_ALL_INNER, lo que exige una ejecución en serie porque toda la tabla interna fue redistribuida a un nodo único. | Hay una vista DS_DIST_ALL_INNER para la combinación hash en el plan de consulta. | Revise la elección de la estrategia de distribución para distribuir la tabla interna, en lugar de la externa. | Consulte Distribución de datos poco óptima. |

## Análisis del plan de consulta

Ejecute el comando EXPLAIN para obtener un plan de consulta.

Antes de analizar el plan de consulta, deberá estar familiarizado con la lectura del plan. Si desconoce cómo se lee un plan de consulta, le recomendamos leer Creación e interpretación de un plan de consultas antes de continuar.

Pasos para analizar los datos proporcionados por el plan de consulta

1. Identifique los pasos que tienen el costo más elevado. Céntrese en optimizar esos pasos cuando continúe con los demás pasos.
2. Analice los tipos de combinaciones:
   - Bucle anidado: estas combinaciones suelen producirse cuando se omite una condición de combinación. Para conocer las soluciones recomendadas, consulte Bucle anidado.

- Hash y hash join: se utilizan para combinar tablas en las que las columnas de combinación no son claves de distribución ni tampoco claves de ordenación. Para conocer las soluciones recomendadas, consulte Combinación hash.
- Merge join: no es necesario realizar cambios.

3. Tenga en cuenta qué tabla se utiliza para la combinación interna y cuál para la combinación externa. El motor de consultas, por lo general, selecciona la tabla más pequeña para la combinación interna y la tabla más grande para la combinación externa. Si no ocurre dicha elección, es probable que sus estadísticas estén desactualizadas. Para conocer las soluciones recomendadas, consulte Faltan estadísticas de tablas o están desactualizadas.
4. Analice si hay algunas operaciones de ordenación costosas. En caso afirmativo, consulte Filas desordenadas o mal ordenadas para conocer las soluciones recomendadas.
5. Busque los siguientes operadores de difusión donde haya operaciones costosas:
   - DS_BCAST_INNER: indica que la tabla se transmite a todos los nodos de computación. Esto está bien para una tabla pequeña, pero no es ideal para una tabla más grande.
   - DS_DIST_ALL_INNER: indica que toda la carga de trabajo está en un único sector.
   - DS_DIST_BOTH: indica una gran redistribución.

   Para conocer las soluciones recomendadas para estas situaciones, consulte Distribución de datos poco óptima.

## Análisis del resumen de consultas

Para obtener información más detallada sobre las estadísticas y los pasos de ejecución que la que proporciona el plan de consulta generado por el comando EXPLAIN, utilice las vistas de sistema SVL_QUERY_SUMMARY y SVL_QUERY_REPORT.

SVL_QUERY_SUMMARY proporciona estadísticas de consultas por secuencia. Puede utilizar la información que proporciona para identificar problemas con pasos costosos, pasos que exigen mucho tiempo de ejecución y pasos que se guardan en el disco.

La vista de sistema SVL_QUERY_REPORT le permite ver información similar a la vista SVL_QUERY_SUMMARY, solo que por sector de nodo de computación, en lugar de por secuencia. Puede utilizar la información de nivel de sector para detectar una distribución irregular en el clúster (también denominada sesgo en la distribución de datos), lo cual exige que algunos nodos trabajen más que otros y perjudica el rendimiento de las consultas.

Temas

- Uso de la vista SVL_QUERY_SUMMARY
- Uso de la vista SVL_QUERY_REPORT
- Mapeo entre el plan de consulta y el resumen de la consulta

## Uso de la vista SVL_QUERY_SUMMARY

Para analizar la información resumida de una consulta por secuencia con SVL_QUERY_SUMMARY, haga lo siguiente:

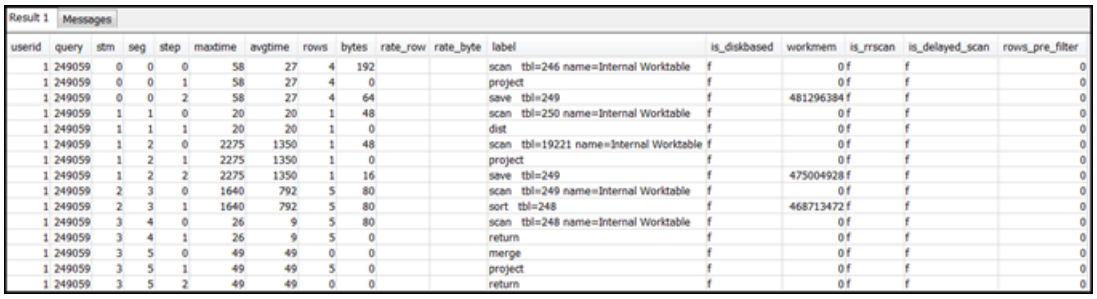
1. Ejecute la siguiente consulta para determinar el ID de su consulta:

   ```
   select query, elapsed, substring
   from svl_qlog
   order by query
   desc limit 5;
   ```

   Examine el texto truncado de la consulta en el campo `substring` para determinar qué valor de `query` representa su consulta. Si ejecutó la consulta más de una vez, utilice el valor `query` de la fila con el valor de `elapsed` más bajo. Esa es la fila de la versión compilada. Si ha ejecutado distintas consultas, puede aumentar el valor que utilizó la cláusula LIMIT para asegurarse de que su consulta esté incluida.
2. Seleccione las filas de SVL_QUERY_SUMMARY para su consulta. Ordene los resultados por secuencia, segmento y paso:

   ```
   select * from svl_query_summary where query = MyQueryID order by stm, seg, step;
   ```

   A continuación se muestra un resultado de ejemplo.



| userid | query | stm | seg | step | maxtime | avgtime | rows | bytes | rate_row | rate_byte | label | is_diskbased | workmem | is_rrscan | is_delayed_scan | rows_pre_filter |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 249059 | 0 | 0 | 0 | 58 | 27 | 4 | 192 | | | scan tbl=246 name=Internal Worktable | f | 0 | f | f | 0 |
| 1 | 249059 | 0 | 0 | 1 | 58 | 27 | 4 | 0 | | | project | f | 0 | f | f | 0 |
| 1 | 249059 | 0 | 0 | 2 | 58 | 27 | 4 | 64 | | | save tbl=249 | f | 481296384 | f | f | 0 |
| 1 | 249059 | 1 | 1 | 0 | 20 | 20 | 1 | 48 | | | scan tbl=250 name=Internal Worktable | f | 0 | f | f | 0 |
| 1 | 249059 | 1 | 1 | 1 | 20 | 20 | 1 | 0 | | | dist | f | 0 | f | f | 0 |
| 1 | 249059 | 1 | 2 | 0 | 2275 | 1350 | 1 | 48 | | | scan tbl=19221 name=Internal Worktable | f | 0 | f | f | 0 |
| 1 | 249059 | 1 | 2 | 1 | 2275 | 1350 | 1 | 0 | | | project | f | 0 | f | f | 0 |
| 1 | 249059 | 1 | 2 | 2 | 2275 | 1350 | 1 | 16 | | | save tbl=249 | f | 475004928 | f | f | 0 |
| 1 | 249059 | 2 | 3 | 0 | 1640 | 792 | 5 | 80 | | | scan tbl=249 name=Internal Worktable | f | 0 | f | f | 0 |
| 1 | 249059 | 2 | 3 | 1 | 1640 | 792 | 5 | 80 | | | sort tbl=248 | f | 468713472 | f | f | 0 |
| 1 | 249059 | 3 | 4 | 0 | 26 | 9 | 5 | 80 | | | scan tbl=248 name=Internal Worktable | f | 0 | f | f | 0 |
| 1 | 249059 | 3 | 4 | 1 | 26 | 9 | 5 | 0 | | | return | f | 0 | f | f | 0 |
| 1 | 249059 | 3 | 5 | 0 | 49 | 49 | 0 | 0 | | | merge | f | 0 | f | f | 0 |
| 1 | 249059 | 3 | 5 | 1 | 49 | 49 | 5 | 0 | | | project | f | 0 | f | f | 0 |
| 1 | 249059 | 3 | 5 | 2 | 49 | 49 | 0 | 0 | | | return | f | 0 | f | f | 0 |

3. Haga corresponder los pasos con las operaciones del plan de consulta utilizando la información en Mapeo entre el plan de consulta y el resumen de la consulta. Tienen que tener, aproximadamente,

los mismos valores de filas y de bytes (filas * ancho del plan de consulta). En caso de no coincidir, consulte Faltan estadísticas de tablas o están desactualizadas para conocer las soluciones recomendadas.

4. Controle si el campo `is_diskbased` tiene el valor `t` ("true") para algún paso. Hash, aggregate y sort son los operadores que, probablemente, se guarden en el disco si el sistema no tiene suficiente memoria asignada para el procesamiento de consultas.

   Si `is_diskbased` tiene el valor "true", consulte Memoria insuficiente asignada a la consulta para conocer las soluciones recomendadas.

5. Controle los valores del campo `label` y vea si hay una secuencia AGG-DIST-AGG en alguna parte de los pasos. Si dicha secuencia está presente, esto indica que hay una agregación de dos pasos, que es costosa. Para reparar esto, cambie la cláusula GROUP BY para utilizar la clave de distribución (la primera clave, si hay varias).

6. Revise el valor de `maxtime` de cada segmento (es el mismo en todos los pasos del segmento). Identifique el segmento con el valor más alto de `maxtime` y revise los pasos en este segmento para los siguientes operadores.

   > **Note**
   >
   > Un valor elevado de `maxtime` no indica necesariamente un problema con el segmento. Más allá de que el valor sea elevado, es posible que el segmento no haya tardado mucho tiempo en procesarse. Todos los segmentos de una secuencia comienzan a programarse al mismo tiempo. Sin embargo, es posible que algunos segmentos posteriores no se puedan ejecutar hasta obtener datos de los segmentos anteriores. Es posible que este efecto haga parecer que hayan tardado mucho tiempo porque su valor de `maxtime` incluye tanto su tiempo de espera como su tiempo de procesamiento.

   - BCAST o DIST: en estos casos, el valor elevado de `maxtime` puede resultar de la redistribución de una gran cantidad de filas. Para conocer las soluciones recomendadas, consulte Distribución de datos poco óptima.
   - HJOIN (hash join): si el paso en cuestión tiene un valor muy elevado en el campo `rows` en comparación con el valor de `rows` en el paso final RETURN de la consulta, consulte Combinación hash para conocer las soluciones recomendadas.

- SCAN/SORT: busque una secuencia SCAN, SORT, SCAN, MERGE de pasos justo antes de un paso de combinación. Este patrón indica que los datos desordenados se examinan, ordenan y, luego, fusionan con el área ordenada de la tabla.

  Advierta si el valor de rows del paso SCAN tiene un valor muy elevado en comparación con el valor de rows del paso final RETURN de la consulta. Este patrón indica que el motor de ejecución está examinando filas que, luego, se descartan, lo cual es poco eficiente. Para conocer las soluciones recomendadas, consulte Predicado poco restrictivo.

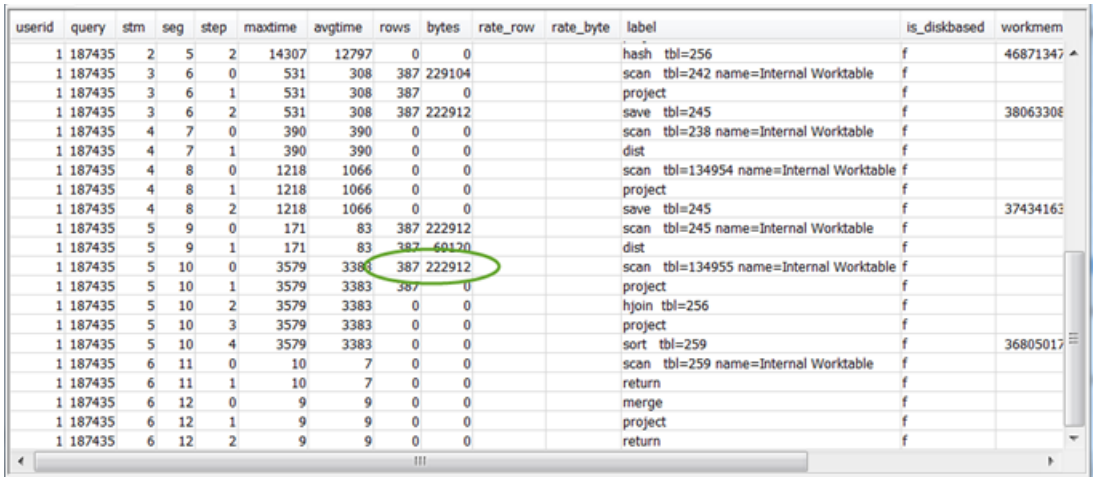
  Si el valor de `maxtime` del paso SCAN es elevado, consulte Cláusula WHERE poco óptima para conocer las soluciones recomendadas.

  Si el valor de `rows` del paso SORT no es cero, consulte Filas desordenadas o mal ordenadas para conocer las soluciones recomendadas.

7. Revise los valores de `rows` y de `bytes` de los pasos 5 a 10 anteriores al paso final RETURN para hacerse una idea de la cantidad de datos que se devuelven al cliente. Este proceso puede ser un arte en sí mismo.

   Por ejemplo, en el siguiente resumen de consulta de ejemplo, el tercer paso PROJECT proporciona un valor de `rows`, pero no un valor de `bytes`. Si analiza los pasos anteriores para buscar uno con el mismo valor de `rows`, encontrará que el paso SCAN proporciona información relacionada con filas y bytes.

   Lo que sigue es un ejemplo de resultado.



| userid | query | stm | seg | step | maxtime | avgtime | rows | bytes | rate_row | rate_byte | label | is_diskbased | workmem |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 187435 | 2 | 5 | 2 | 14307 | 12797 | 0 | 0 | | | hash tbl=256 | f | 46871347 |
| 1 | 187435 | 3 | 6 | 0 | 531 | 308 | 387 | 229104 | | | scan tbl=242 name=Internal Worktable | f | |
| 1 | 187435 | 3 | 6 | 1 | 531 | 308 | 387 | 0 | | | project | f | |
| 1 | 187435 | 3 | 6 | 2 | 531 | 308 | 387 | 222912 | | | save tbl=245 | f | 38063308 |
| 1 | 187435 | 4 | 7 | 0 | 390 | 390 | 0 | 0 | | | scan tbl=238 name=Internal Worktable | f | |
| 1 | 187435 | 4 | 7 | 1 | 390 | 390 | 0 | 0 | | | dist | f | |
| 1 | 187435 | 4 | 8 | 0 | 1218 | 1066 | 0 | 0 | | | scan tbl=134954 name=Internal Worktable | f | |
| 1 | 187435 | 4 | 8 | 1 | 1218 | 1066 | 0 | 0 | | | project | f | |
| 1 | 187435 | 4 | 8 | 2 | 1218 | 1066 | 0 | 0 | | | save tbl=245 | f | 37434163 |
| 1 | 187435 | 5 | 9 | 0 | 171 | 83 | 387 | 222912 | | | scan tbl=245 name=Internal Worktable | f | |
| 1 | 187435 | 5 | 9 | 1 | 171 | 83 | 387 | 60120 | | | dist | f | |
| 1 | 187435 | 5 | 10 | 0 | 3579 | 3383 | 387 | 222912 | | | scan tbl=134955 name=Internal Worktable | f | |
| 1 | 187435 | 5 | 10 | 1 | 3579 | 3383 | 387 | 0 | | | project | f | |
| 1 | 187435 | 5 | 10 | 2 | 3579 | 3383 | 0 | 0 | | | hjoin tbl=256 | f | |
| 1 | 187435 | 5 | 10 | 3 | 3579 | 3383 | 0 | 0 | | | project | f | |
| 1 | 187435 | 5 | 10 | 4 | 3579 | 3383 | 0 | 0 | | | sort tbl=259 | f | 36805017 |
| 1 | 187435 | 6 | 11 | 0 | 10 | 7 | 0 | 0 | | | scan tbl=259 name=Internal Worktable | f | |
| 1 | 187435 | 6 | 11 | 1 | 10 | 7 | 0 | 0 | | | return | f | |
| 1 | 187435 | 6 | 12 | 0 | 9 | 9 | 0 | 0 | | | merge | f | |
| 1 | 187435 | 6 | 12 | 1 | 9 | 9 | 0 | 0 | | | project | f | |
| 1 | 187435 | 6 | 12 | 2 | 9 | 9 | 0 | 0 | | | return | f | |

   Si devuelve un volumen de datos excepcionalmente grande, consulte Conjunto de resultados muy grande para conocer las soluciones recomendadas.

8. Vea si el valor de `bytes` es elevado en comparación con el valor de `rows` para cualquier paso, en relación con los demás pasos. Este patrón puede indicar que está seleccionando demasiadas columnas. Para conocer las soluciones recomendadas, consulte Lista SELECT grande.

## Uso de la vista SVL_QUERY_REPORT

Para analizar la información resumida de una consulta por sector con SVL_QUERY_REPORT, haga lo siguiente:

1. Ejecute la siguiente función para determinar el ID de su consulta:

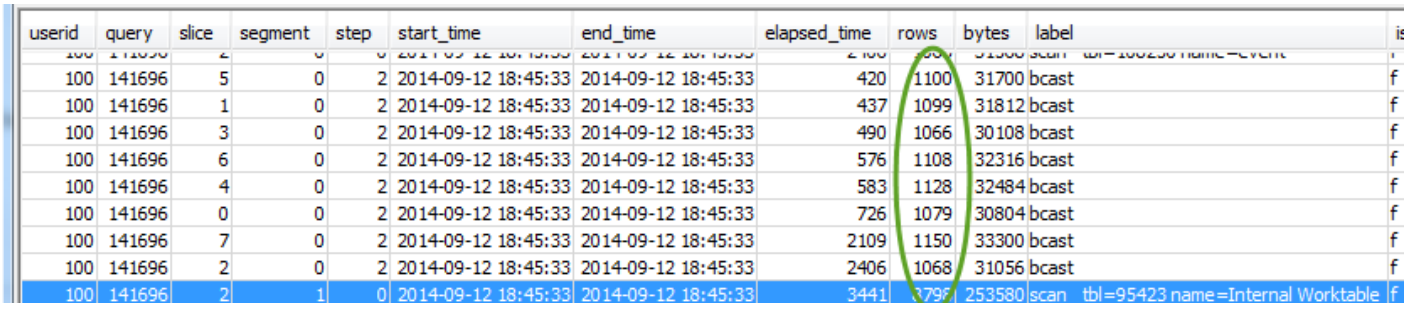
```
select query, elapsed, substring
from svl_qlog
order by query
desc limit 5;
```

   Examine el texto truncado de la consulta en el campo `substring` para determinar qué valor de `query` representa su consulta. Si ejecutó la consulta más de una vez, utilice el valor `query` de la fila con el valor de `elapsed` más bajo. Esa es la fila de la versión compilada. Si ha ejecutado distintas consultas, puede aumentar el valor que utilizó la cláusula LIMIT para asegurarse de que su consulta esté incluida.

2. Seleccione las filas de SVL_QUERY_REPORT para su consulta. Ordene los resultados por secuencia, paso, tiempo transcurrido y filas:

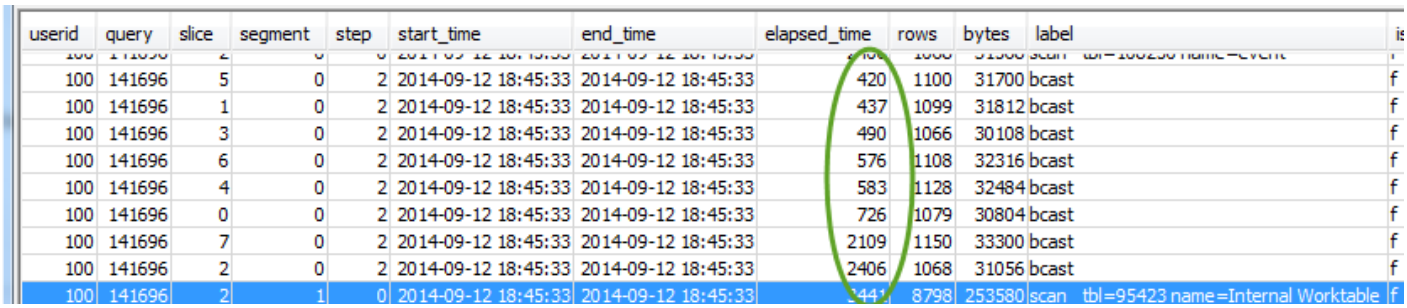
```
select * from svl_query_report where query = MyQueryID order by segment, step,
 elapsed_time, rows;
```

3. En cada paso, controle que todos los sectores estén procesando, aproximadamente, la misma cantidad de filas:



| userid | query | slice | segment | step | start_time | end_time | elapsed_time | rows | bytes | label | is |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 141696 | 5 | 0 | 2 | 2014-09-12 18:45:33 | 2014-09-12 18:45:33 | 420 | 1100 | 31700 | bcast | f |
| 100 | 141696 | 1 | 0 | 2 | 2014-09-12 18:45:33 | 2014-09-12 18:45:33 | 437 | 1099 | 31812 | bcast | f |
| 100 | 141696 | 3 | 0 | 2 | 2014-09-12 18:45:33 | 2014-09-12 18:45:33 | 490 | 1066 | 30108 | bcast | f |
| 100 | 141696 | 6 | 0 | 2 | 2014-09-12 18:45:33 | 2014-09-12 18:45:33 | 576 | 1108 | 32316 | bcast | f |
| 100 | 141696 | 4 | 0 | 2 | 2014-09-12 18:45:33 | 2014-09-12 18:45:33 | 583 | 1128 | 32484 | bcast | f |
| 100 | 141696 | 0 | 0 | 2 | 2014-09-12 18:45:33 | 2014-09-12 18:45:33 | 726 | 1079 | 30804 | bcast | f |
| 100 | 141696 | 7 | 0 | 2 | 2014-09-12 18:45:33 | 2014-09-12 18:45:33 | 2109 | 1150 | 33300 | bcast | f |
| 100 | 141696 | 2 | 0 | 2 | 2014-09-12 18:45:33 | 2014-09-12 18:45:33 | 2406 | 1068 | 31056 | bcast | f |
| 100 | 141696 | 2 | 1 | 0 | 2014-09-12 18:45:33 | 2014-09-12 18:45:33 | 3441 | 3798 | 253580 | scan tbl=95423 name=Internal Worktable | f |

Además, controle que a todos los sectores les tome, aproximadamente, la misma cantidad de tiempo:



| userid | query | slice | segment | step | start_time | end_time | elapsed_time | rows | bytes | label | is |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 141696 | 5 | 0 | 2 | 2014-09-12 18:45:33 | 2014-09-12 18:45:33 | 420 | 1100 | 31700 | bcast | f |
| 100 | 141696 | 1 | 0 | 2 | 2014-09-12 18:45:33 | 2014-09-12 18:45:33 | 437 | 1099 | 31812 | bcast | f |
| 100 | 141696 | 3 | 0 | 2 | 2014-09-12 18:45:33 | 2014-09-12 18:45:33 | 490 | 1066 | 30108 | bcast | f |
| 100 | 141696 | 6 | 0 | 2 | 2014-09-12 18:45:33 | 2014-09-12 18:45:33 | 576 | 1108 | 32316 | bcast | f |
| 100 | 141696 | 4 | 0 | 2 | 2014-09-12 18:45:33 | 2014-09-12 18:45:33 | 583 | 1128 | 32484 | bcast | f |
| 100 | 141696 | 0 | 0 | 2 | 2014-09-12 18:45:33 | 2014-09-12 18:45:33 | 726 | 1079 | 30804 | bcast | f |
| 100 | 141696 | 7 | 0 | 2 | 2014-09-12 18:45:33 | 2014-09-12 18:45:33 | 2109 | 1150 | 33300 | bcast | f |
| 100 | 141696 | 2 | 0 | 2 | 2014-09-12 18:45:33 | 2014-09-12 18:45:33 | 2406 | 1068 | 31056 | bcast | f |
| 100 | 141696 | 2 | 1 | 0 | 2014-09-12 18:45:33 | 2014-09-12 18:45:33 | 2441 | 8798 | 253580 | scan tbl=95423 name=Internal Worktable | f |

Si hay grandes diferencias entre estos valores, puede indicar que hay un sesgo en la distribución de datos debido a un estilo de distribución que no es el óptimo para esta consulta en particular. Para conocer las soluciones recomendadas, consulte Distribución de datos poco óptima.

## Mapeo entre el plan de consulta y el resumen de la consulta

Al analizar el resumen de consultas, puede obtener más información mediante la asignación de las operaciones desde el plan de consulta hasta los pasos (identificados por los valores del campo de etiqueta) en el resumen de consultas. La siguiente tabla asigna las operaciones del plan de consulta a los pasos del resumen de consultas.

| Operación del plan de consulta | Valor del campo Label | Descripción |
|---|---|---|
| Agregado<br>HashAggregate<br>GroupAggregate | AGGR | Evalúa las funciones de agregación y las condiciones GROUP BY. |
| DS_BCAST_INNER | BCAST (difusión) | Difunde una tabla completa o un conjunto de filas (por ejemplo, un conjunto de filas filtrado de una tabla) a todos los nodos. |

| Operación del plan de consulta | Valor del campo Label | Descripción |
|---|---|---|
| No figura en el plan de consulta | DELETE | Elimina filas de tablas. |
| DS_DIST_NONE<br>DS_DIST_ALL_NONE<br>DS_DIST_INNER<br>DS_DIST_ALL_INNER<br>DS_DIST_ALL_BOTH | DIST (distribuir) | Distribuye las filas en nodos para hacer combinaciones y otros procesamientos en paralelo. |
| HASH | HASH | Construye tablas hash para utilizarlas en las combinaci ones hash. |
| Hash Join | HJOIN (combinación hash) | Ejecuta una unión hash de dos tablas o de conjuntos de resultados intermedios. |
| No figura en el plan de consulta | INSERT | Inserta filas en tablas. |
| Límite | LIMIT | Aplica una cláusula LIMIT a los conjuntos de resultados. |
| Merge | MERGE | Fusiona filas que surgen de operaciones de ordenación o de combinación en paralelo. |
| Merge Join | MJOIN (combinación de fusión) | Ejecuta una unión de fusión de dos tablas o de conjuntos de resultados intermedios. |

| Operación del plan de consulta | Valor del campo Label | Descripción |
|---|---|---|
| Nested Loop | NLOOP (bucle anidado) | Ejecuta una unión de bucle anidado de dos tablas o de conjuntos de resultados intermedios. |
| No figura en el plan de consulta | PARSE | Analiza cadenas en valores binarios para la carga. |
| Proyecto | PROJECT | Evalúa expresiones. |
| Network | RETURN | Devuelve filas al nodo principal o al cliente. |
| No figura en el plan de consulta | SAVE | Materializa las filas para utilizarlas en el próximo paso de procesamiento. |
| Seq Scan | SCAN | Examina tablas o conjuntos de resultados intermedios. |
| Sort | SORT | Ordena filas o conjuntos de resultados intermedios tal como lo exigen otras operaciones posteriores (como combinaciones o agregaciones) o para cumplir con una cláusula ORDER BY. |
| Unique | UNIQUE | Aplica una cláusula SELECT DISTINCT o elimina los duplicados según lo exijan otras operaciones. |
| Ventana | WINDOW | Calcula la adición y la clasifica ción de funciones de ventana. |

# Mejora del rendimiento de las consultas

A continuación, se describen algunos problemas frecuentes que afectan al rendimiento de las consultas de Amazon Redshift y se detallan instrucciones acerca de cómo diagnosticarlos y resolverlos.

Temas


- Faltan estadísticas de tablas o están desactualizadas
- Bucle anidado
- Combinación hash
- Filas fantasma o filas sin confirmar
- Filas desordenadas o mal ordenadas
- Distribución de datos poco óptima
- Memoria insuficiente asignada a la consulta
- Cláusula WHERE poco óptima
- Predicado poco restrictivo
- Conjunto de resultados muy grande
- Lista SELECT grande


## Faltan estadísticas de tablas o están desactualizadas

Si faltan estadísticas de tablas o si estas están desactualizadas, es posible que vea lo siguiente:

- Un mensaje de advertencia en los resultados del comando EXPLAIN.
- Un evento de alerta en STL_ALERT_EVENT_LOG que indica que faltan estadísticas. Para obtener más información, consulte Revisión de alertas de consultas.

Para reparar este problema, ejecute el comando ANALYZE.

## Bucle anidado

Si hay un bucle anidado, es posible que vea un evento de alerta de bucle anidado en STL_ALERT_EVENT_LOG. También puede identificar este tipo de evento si ejecuta la consulta en Identificación de consultas con bucles anidados. Para obtener más información, consulte Revisión de alertas de consultas.

Para reparar esto, asegúrese de que su consulta no tenga combinaciones cruzadas o elimínelas de ser posible. Las combinaciones cruzadas son combinaciones sin una condición de combinación que dan lugar a un producto cartesiano de dos tablas. Por lo general, se ejecutan como uniones de bucles anidados, que son los tipos de unión posibles más lentos.

### Combinación hash

Si hay una combinación hash, es posible que vea lo siguiente:

- Operaciones hash y hash join en el plan de consulta. Para obtener más información, consulte Análisis del plan de consulta.
- Un paso HJOIN en el segmento con el valor máximo de maxtime en SVL_QUERY_SUMMARY. Para obtener más información, consulte Uso de la vista SVL_QUERY_SUMMARY.

Para reparar este problema, puede optar por varios métodos:

- Reescribir la consulta para usar una combinación de fusión de ser posible. Puede hacerlo si especifica cuáles son las columnas de combinación que son claves de distribución y claves de ordenación al mismo tiempo.
- Si el paso HJOIN en SVL_QUERY_SUMMARY tiene un valor muy elevado en el campo rows en comparación con el valor de rows en el paso final RETURN de la consulta, compruebe si puede volver a escribir la consulta para combinarla en una única columna. Cuando una consulta no se combina en una única columna, como en el caso de una clave principal, aumenta la cantidad de filas involucradas en la combinación.

### Filas fantasma o filas sin confirmar

Si hay filas fantasma o filas sin confirmar, es posible que vea un evento de alerta en STL_ALERT_EVENT_LOG que indique que hay demasiadas filas fantasma. Para obtener más información, consulte Revisión de alertas de consultas.

Para reparar este problema, puede optar por varios métodos:

- Revise la pestaña Loads (Cargas) de la consola de Amazon Redshift para ver si hay operaciones de carga activas en cualquiera de las tablas de consultas. Si ve operaciones de carga activas, espere a que se completen antes de realizar cualquier medida.
- Si no hay operaciones de carga activas, ejecute el comando VACUUM en las tablas de la consulta para quitar las filas eliminadas.

## Filas desordenadas o mal ordenadas

Si hay filas desordenadas o mal ordenadas, es posible que vea un evento de alerta de filtro muy selectivo en STL_ALERT_EVENT_LOG. Para obtener más información, consulte Revisión de alertas de consultas.

También puede controlar si alguna de las tablas de su consulta tiene grandes áreas desordenadas si ejecuta la consulta en Identificación de tablas con sesgo de datos o con filas desordenadas.

Para reparar este problema, puede optar por varios métodos:

- Ejecute el comando VACUUM en las tablas de la consulta para reordenar las filas.
- Revise las claves de ordenación en las tablas de consulta para ver si puede hacer alguna mejora. Recuerde comparar el rendimiento de esta consulta con el rendimiento de otras consultas importantes y con el sistema en general antes de implementar cualquier cambio. Para obtener más información, consulte Claves de clasificación.

## Distribución de datos poco óptima

Si la distribución de datos no es la óptima, es posible que vea lo siguiente:

- Aparecerá un evento de alerta de ejecución en serie, difusión o distribución de gran tamaño en STL_ALERT_EVENT_LOG. Para obtener más información, consulte Revisión de alertas de consultas.
- Los sectores no están procesando aproximadamente la misma cantidad de filas en un paso determinado. Para obtener más información, consulte Uso de la vista SVL_QUERY_REPORT.
- Los sectores no están tardando aproximadamente el mismo tiempo en un paso determinado. Para obtener más información, consulte Uso de la vista SVL_QUERY_REPORT.

Si ninguna de las opciones anteriores es verdadera, también puede controlar si alguna de las tablas de su consulta tiene un sesgo en sus datos mediante la ejecución de la consulta en Identificación de tablas con sesgo de datos o con filas desordenadas.

Para solucionar este problema, consulte los estilos de distribución de las tablas de la consulta y determine si se puede hacer alguna mejora. Recuerde comparar el rendimiento de esta consulta con el rendimiento de otras consultas importantes y con el sistema en general antes de implementar cualquier cambio. Para obtener más información, consulte Distribución de datos para la optimización de consultas.

## Memoria insuficiente asignada a la consulta

Si la memoria asignada a su consulta es insuficiente, es probable que vea un paso en SVL_QUERY_SUMMARY que tenga un valor "true" en `is_diskbased`. Para obtener más información, consulte Uso de la vista SVL_QUERY_SUMMARY.

Para reparar este problema, asigne más memoria a la consulta aumentando temporalmente la cantidad de slots que usa la consulta. Workload Management (WLM, Administración de cargas de trabajo) reserva slots en una cola de consultas equivalente al nivel de simultaneidad definido para la cola. Por ejemplo, una cola con un nivel de simultaneidad de 5 tiene 5 slots. La memoria asignada a la cola se asigna en partes iguales a cada slot. Asignar muchos slots a una consulta le proporciona a esa consulta acceso a la memoria de todos esos slots. Para obtener más información acerca de cómo aumentar temporalmente los slots para una consulta, consulte wlm_query_slot_count.

## Cláusula WHERE poco óptima

Si su cláusula WHERE genera demasiados exámenes de tablas, es posible que vea un paso SCAN en el segmento con el valor máximo de `maxtime` en SVL_QUERY_SUMMARY. Para obtener más información, consulte Uso de la vista SVL_QUERY_SUMMARY.

Para reparar este problema, añada una cláusula WHERE a la consulta basada en la columna de ordenación principal de la tabla más grande. Este enfoque ayuda a reducir el tiempo de análisis. Para obtener más información, consulte Prácticas recomendadas de Amazon Redshift para el diseño de tablas.

## Predicado poco restrictivo

Si su consulta tiene un predicado poco restrictivo, es posible que vea un paso SCAN en el segmento con el valor más alto de `maxtime` en SVL_QUERY_SUMMARY que tenga un valor de `rows` muy elevado, en comparación con el valor de `rows` en el paso final RETURN de la consulta. Para obtener más información, consulte Uso de la vista SVL_QUERY_SUMMARY.

Para reparar este problema, añada un predicado a la consulta o haga que el predicado existente sea más restrictivo para acotar la salida.

## Conjunto de resultados muy grande

Si la consulta devuelve un conjunto de resultados muy grande, considere la posibilidad de reescribirla de forma que utilice UNLOAD para escribir los resultados en Amazon S3. Este método mejora el rendimiento del paso RETURN, ya que aprovecha el procesamiento en paralelo. Para

obtener más información acerca de cómo controlar si hay conjuntos de resultados muy grandes, consulte Uso de la vista SVL_QUERY_SUMMARY.

### Lista SELECT grande

Si su consulta tiene una lista SELECT excepcionalmente grande, es probable que vea un valor de `bytes` elevado, en comparación con el valor de `rows` en cualquier paso (con relación a otros pasos) en SVL_QUERY_SUMMARY. Este valor elevado de `bytes` puede indicar que está seleccionando muchas columnas. Para obtener más información, consulte Uso de la vista SVL_QUERY_SUMMARY.

Para reparar este problema, revise las columnas que está seleccionando y vea si se pueden eliminar.

## Consultas de diagnóstico para el ajuste de consultas

Utilice las consultas a continuación para identificar problemas con las consultas o con sus tablas subyacentes que pueden afectar el rendimiento de las consultas. Le recomendamos usar estas consultas con los procesos de ajuste de consultas descritos en Análisis y mejora de consultas.

> **Note**
>
> Estas consultas son para clústeres aprovisionados de Amazon Redshift. Estas consultas no se deben usar con grupos de trabajo de Redshift sin servidor.

Temas

- Identificación de consultas que deben ajustarse con prioridad
- Identificación de tablas con sesgo de datos o con filas desordenadas
- Identificación de consultas con bucles anidados
- Revisión de los tiempos de espera de las colas de consultas
- Revisión de las alertas de consultas por tablas
- Identificación de tablas con estadísticas faltantes

### Identificación de consultas que deben ajustarse con prioridad

La siguiente consulta identifica las 50 instrucciones que se han ejecutado en los últimos 7 días y que más tiempo demandaron. Puede utilizar los resultados para identificar las consultas que están tardando excesivamente. También puede identificar las consultas que se ejecutan con frecuencia (las

que aparecen más de una vez en el conjunto de resultados). Estas consultas suelen ser prioritarias para realizar ajustes destinados a mejorar el rendimiento del sistema.

Esta consulta también proporciona un recuento de los eventos de alerta asociados a cada consulta identificada. Estas alertas proporcionan detalles que puede utilizar para mejorar el rendimiento de la consulta. Para obtener más información, consulte Revisión de alertas de consultas.

```
select trim(database) as db, count(query) as n_qry,
max(substring (qrytext,1,80)) as qrytext,
min(run_minutes) as "min" ,
max(run_minutes) as "max",
avg(run_minutes) as "avg", sum(run_minutes) as total,
max(query) as max_query_id,
max(starttime)::date as last_run,
sum(alerts) as alerts, aborted
from (select userid, label, stl_query.query,
trim(database) as database,
trim(querytxt) as qrytext,
md5(trim(querytxt)) as qry_md5,
starttime, endtime,
(datediff(seconds, starttime,endtime)::numeric(12,2))/60 as run_minutes,
alrt.num_events as alerts, aborted
from stl_query
left outer join
(select query, 1 as num_events from stl_alert_event_log group by query ) as alrt
on alrt.query = stl_query.query
where userid <> 1 and starttime >=  dateadd(day, -7, current_date))
group by database, label, qry_md5, aborted
order by total desc limit 50;
```

## Identificación de tablas con sesgo de datos o con filas desordenadas

La siguiente consulta identifica las tablas que tienen una distribución irregular de datos (sesgo de datos) o un alto porcentaje de filas desordenadas.

Un valor bajo de `skew` indica que los datos de la tabla están distribuidos correctamente. Si una tabla tiene un valor de `skew` de 4,00 o menor, considere la opción de modificar su estilo de distribución de datos. Para obtener más información, consulte Distribución de datos poco óptima.

Si una tabla tiene un valor de `pct_unsorted` mayor que 20 por ciento, considere la opción de ejecutar el comando VACUUM. Para obtener más información, consulte Filas desordenadas o mal ordenadas.

Revise también los valores `mbytes` y `pct_of_total` de cada tabla. Estas columnas identifican el tamaño de la tabla y el porcentaje de espacio bruto en disco que consume la tabla. El espacio de disco en bruto incluye el espacio que reserva Amazon Redshift para uso interno, por lo que supera su capacidad nominal, la cual corresponde a la cantidad de espacio de disco disponible para el usuario. Utilice esta información para verificar que tiene espacio libre en el disco que equivalga al menos a 2,5 veces el tamaño de su tabla más grande. Si tiene este espacio disponible, el sistema podrá guardar los resultados intermedios en el disco cuando procese consultas complejas.

```
select trim(pgn.nspname) as schema,
trim(a.name) as table, id as tableid,
decode(pgc.reldiststyle,0, 'even',1,det.distkey ,8,'all') as distkey,
 dist_ratio.ratio::decimal(10,4) as skew,
det.head_sort as "sortkey",
det.n_sortkeys as "#sks", b.mbytes,
decode(b.mbytes,0,0,((b.mbytes/part.total::decimal)*100)::decimal(5,2)) as
 pct_of_total,
decode(det.max_enc,0,'n','y') as enc, a.rows,
decode( det.n_sortkeys, 0, null, a.unsorted_rows ) as unsorted_rows ,
decode( det.n_sortkeys, 0, null, decode( a.rows,0,0, (a.unsorted_rows::decimal(32)/
a.rows)*100) )::decimal(5,2) as pct_unsorted
from (select db_id, id, name, sum(rows) as rows,
sum(rows)-sum(sorted_rows) as unsorted_rows
from stv_tbl_perm a
group by db_id, id, name) as a
join pg_class as pgc on pgc.oid = a.id
join pg_namespace as pgn on pgn.oid = pgc.relnamespace
left outer join (select tbl, count(*) as mbytes
from stv_blocklist group by tbl) b on a.id=b.tbl
inner join (select attrelid,
min(case attisdistkey when 't' then attname else null end) as "distkey",
min(case attsortkeyord when 1 then attname  else null end ) as head_sort ,
max(attsortkeyord) as n_sortkeys,
max(attencodingtype) as max_enc
from pg_attribute group by 1) as det
on det.attrelid = a.id
inner join ( select tbl, max(mbytes)::decimal(32)/min(mbytes) as ratio
from (select tbl, trim(name) as name, slice, count(*) as mbytes
from svv_diskusage group by tbl, name, slice )
group by tbl, name ) as dist_ratio on a.id = dist_ratio.tbl
join ( select sum(capacity) as  total
from stv_partitions where part_begin=0 ) as part on 1=1
where mbytes is not null
```

```
order by  mbytes desc;
```

## Identificación de consultas con bucles anidados

La siguiente consulta identifica consultas que han registrado eventos de alerta por bucles anidados. Para obtener más información acerca de cómo reparar una condición de bucle anidado, consulte Bucle anidado.

```
select query, trim(querytxt) as SQL, starttime
from stl_query
where query in (
select distinct query
from stl_alert_event_log
where event like 'Nested Loop Join in the query plan%')
order by starttime desc;
```

## Revisión de los tiempos de espera de las colas de consultas

La siguiente consulta muestra cuánto tiempo han esperado las consultas recientes hasta que se abriera un lugar en una cola de consultas para poder ejecutarse. Si advierte una tendencia de tiempos de espera prolongados, puede modificar la configuración de colas de consultas para obtener un mejor rendimiento. Para obtener más información, consulte Implementación de WLM manual.

```
select trim(database) as DB , w.query,
substring(q.querytxt, 1, 100) as querytxt,  w.queue_start_time,
w.service_class as class, w.slot_count as slots,
w.total_queue_time/1000000 as queue_seconds,
w.total_exec_time/1000000 exec_seconds, (w.total_queue_time+w.total_Exec_time)/1000000
 as total_seconds
from stl_wlm_query w
left join stl_query q on q.query = w.query and q.userid = w.userid
where w.queue_start_Time >= dateadd(day, -7, current_Date)
and w.total_queue_Time > 0  and w.userid >1
and q.starttime >= dateadd(day, -7, current_Date)
order by w.total_queue_time desc, w.queue_start_time desc limit 35;
```

## Revisión de las alertas de consultas por tablas

La siguiente consulta identifica las tablas que han registrado eventos de alerta y también identifica qué tipo de alertas se activan con mayor frecuencia.

Si el valor de `minutes` de una fila con una tabla identificada es elevado, controle si esa tabla necesita un mantenimiento de rutina, como tener que ejecutar ANALYZE o VACUUM en esa tabla.

Si el valor `count` es elevado para una fila, pero el valor de `table` es nulo, ejecute una consulta con STL_ALERT_EVENT_LOG para el valor de `event` asociado a fin de investigar por qué esa alerta se activa con tanta frecuencia.

```
select trim(s.perm_table_name) as table,
(sum(abs(datediff(seconds, s.starttime, s.endtime)))/60)::numeric(24,0) as minutes,
 trim(split_part(l.event,':',1)) as event,  trim(l.solution) as solution,
max(l.query) as sample_query, count(*)
from stl_alert_event_log as l
left join stl_scan as s on s.query = l.query and s.slice = l.slice
and s.segment = l.segment and s.step = l.step
where l.event_time >=  dateadd(day, -7, current_Date)
group by 1,3,4
order by 2 desc,6 desc;
```

### Identificación de tablas con estadísticas faltantes

La siguiente consulta proporciona un recuento de las consultas que está ejecutando con tablas a las cuales le faltan estadísticas. Si esta consulta devuelve alguna fila, fíjese en el valor de `plannode` para determinar cuál es la tabla afectada y, luego, ejecute el comando ANALYZE en esa tabla.

```
select substring(trim(plannode),1,100) as plannode, count(*)
from stl_explain
where plannode like '%missing statistics%'
group by plannode
order by 2 desc;
```

## Solución de problemas de consultas

En esta sección, se proporciona una referencia rápida para identificar y abordar algunos de los problemas más frecuentes y graves que posiblemente enfrente con las consultas de Amazon Redshift.

Temas

- Error en la conexión
- Falta de respuesta de la consulta

- Tiempo de ejecución de consulta demasiado largo
- Error en la carga
- Tiempo de carga excesivo
- Datos de carga incorrectos
- Configuración del parámetro de tamaño de búsqueda de la JDBC

Estas sugerencias le dan una idea inicial para solucionar problemas. También puede consultar los siguientes recursos para obtener información más detallada.

Para obtener información sobre los cambios de comportamiento en la funcionalidad de Amazon Redshift que pueden afectar a la aplicación, consulte Cambios de comportamiento.

- Acceso a las bases de datos y los clústeres de Amazon Redshift
- Optimización de tablas automática
- Carga de datos en Amazon Redshift
- Tutorial: Carga de datos desde Amazon S3

## Error en la conexión

La conexión de consulta puede producir un error por las siguientes razones. Le sugerimos los siguientes enfoques de resolución de problemas.

El cliente no puedes conectarse al servidor

Si está usando certificados SSL o de servidor, primero elimine esta complejidad mientras solucione el problema de conexión. Luego, vuelva a añadir una SSL o certificados del servidor cuando haya encontrado una solución. Para obtener más información, consulte Configuración de las opciones de seguridad para las conexiones en la Guía de administración de Amazon Redshift.

Conexión rechazada

Por lo general, cuando recibe un mensaje de error que le indica que hay un error para establecer una conexión significa que hay un problema con el permiso para obtener acceso al clúster. Para

obtener más información, consulte Errores o rechazos en la conexión en la Guía de administración de Amazon Redshift.

## Falta de respuesta de la consulta

La consulta puede bloquearse o dejar de responder por los siguientes motivos. Le sugerimos los siguientes enfoques de resolución de problemas.

Conexión interrumpida con la base de datos

Reduzca el tamaño de la Maximum Transmission Unit (MTU, Unidad de transmisión máxima). El tamaño de la MTU determina el tamaño máximo, en bytes, de un paquete que puede ser transferido en una trama Ethernet desde su conexión de red. Para obtener más información, consulte Conexión con la base de datos interrumpida en la Guía de administración de Amazon Redshift.

Tiempo de espera de la conexión con la base de datos agotado

Su conexión cliente a la base de datos parece que dejó de funcionar o que superó el tiempo de espera mientras ejecutaba consultas largas, como un comando COPY. En este caso, es posible que observe que la consola de Amazon Redshift muestra que se ha completado la consulta, pero, aparentemente, la propia herramienta cliente sigue ejecutando la consulta. Los resultados de la consulta podrían ser que falta procesar o que está incompleta según cuándo se haya detenido la conexión. Este efecto ocurre cuando un componente intermediario de la red detiene las conexiones inactivas. Para obtener más información, consulte Problema con el tiempo de espera del firewall en la Guía de administración de Amazon Redshift.

Error de memoria insuficiente del cliente con ODBC

Si su aplicación cliente utiliza una conexión Open Database Connectivity (ODBC, Conectividad de base de datos abierta) y su consulta crea un conjunto de resultados que es demasiado grande para entrar en la memoria, puede transmitir el conjunto de resultados a su aplicación cliente por medio del cursor. Para obtener más información, consulte DECLARE y Consideraciones acerca del rendimiento cuando se utilizan cursores.

Error de memoria insuficiente del cliente con JDBC

Cuando intenta recuperar grandes conjuntos de resultados con una conexión Java Database Connectivity (JDBC, Conectividad de base de datos Java), es posible que haya errores de memoria insuficiente del cliente. Para obtener más información, consulte Configuración del parámetro de tamaño de búsqueda de la JDBC.

Hay un posible bloqueo

Si hay un posible bloqueo, pruebe lo siguiente:

- Vea las tablas de sistema STV_LOCKS y STL_TR_CONFLICT para encontrar conflictos relacionados con las actualizaciones de más de una tabla.
- Utilice la función PG_CANCEL_BACKEND para cancelar una o más consultas en conflicto.
- Utilice la función PG_TERMINATE_BACKEND para terminar una sesión, lo que fuerza que las transacciones en ejecución de la sesión terminada liberen todos los bloqueos y reviertan la transacción.
- Programe detenidamente las operaciones simultáneas de escritura. Para obtener más información, consulte Administración de operaciones de escritura simultáneas.

## Tiempo de ejecución de consulta demasiado largo

La consulta puede tardar demasiado por los siguientes motivos. Le sugerimos los siguientes enfoques de resolución de problemas.

Falta de optimización en las tablas

Configure la clave de ordenación, el estilo de distribución y la codificación de compresión de las tablas para sacar el máximo provecho del procesamiento en paralelo. Para obtener más información, consulte Optimización de tablas automática

Consulta guardada en el disco

Es posible que sus consultas se estén guardando en el disco, al menos una parte de la ejecución de las consultas. Para obtener más información, consulte Mejora del rendimiento de las consultas.

Consulta actual es espera hasta que otras consultas finalicen

Puede mejorar el rendimiento general del sistema mediante la creación de colas de consultas y la asignación de diferentes tipos de consultas a las colas correspondientes. Para obtener más información, consulte Administración de la carga de trabajo.

Falta de optimización en las consultas

Analice el plan de explicación para encontrar oportunidades de reescritura de las consultas o de optimización de la base de datos. Para obtener más información, consulte Creación e interpretación de un plan de consultas.

Mayor memoria para la ejecución de consultas

Si una consulta específica necesita más memora, puede aumentar la memoria disponible si aumenta el recuento wlm_query_slot_count.

Ejecución de un comando VACUUM basado en la base de datos

Ejecute el comando VACUUM siempre que añada, elimine o modifique una cantidad importante de filas, a menos que cargue sus datos en orden de clave de ordenación. El comando VACUUM reorganiza sus datos para conservar la ordenación y el rendimiento de la restauración. Para obtener más información, consulte Limpieza de tablas.

## Recursos adicionales para solucionar problemas de consultas de larga duración

A continuación, se indican temas de la vista del sistema y otras secciones de la documentación que resultan útiles para la optimización de consultas:

- La vista del sistema STV_INFLIGHT muestra las consultas que se están ejecutando en el clúster. Puede ser útil utilizarlo junto con STV_RECENTS para determinar qué consultas se están ejecutando en ese momento o se han completado recientemente.
- SYS_QUERY_HISTORY es útil para solucionar problemas. Muestra las consultas DDL y DML con propiedades pertinentes como su estado actual, por ejemplo `running` o `failed`, el tiempo que tardó cada una en ejecutarse y si una consulta se ejecutó en un clúster con escalado de simultaneidad.
- STL_QUERYTEXT captura el texto de la consulta para los comandos SQL. Además, SVV_QUERY_INFLIGHT, que une STL_QUERYTEXT a STV_INFLIGHT, muestra más metadatos de consulta.
- Un conflicto entre transacciones y bloqueos puede ser un posible origen de problemas de rendimiento de las consultas. Para obtener información sobre las transacciones que actualmente mantienen bloqueos en las tablas, consulte SVV_TRANSACTIONS.
- Identificación de consultas que deben ajustarse con prioridad proporciona una consulta de solución de problemas que lo ayuda a determinar qué consultas ejecutadas recientemente consumieron más tiempo. Esto puede ayudarlo a centrar sus esfuerzos en las consultas que necesitan mejoras.
- Si desea profundizar en la administración de consultas y comprender cómo administrar las colas de consultas, Administración de la carga de trabajo le muestra cómo hacerlo. La administración de cargas de trabajo es una característica avanzada y recomendamos la administración automatizada de cargas de trabajo en la mayoría de los casos.

## Error en la carga

La carga de datos puede producir un error por los siguientes motivos. Le sugerimos los siguientes enfoques de resolución de problemas.

Origen de datos en una región de AWS diferente

De manera predeterminada, el bucket de Amazon S3 o la tabla de Amazon DynamoDB especificados en el comando COPY deben estar en la misma región de AWS que el clúster. Si sus datos y su clúster están en regiones diferentes, recibirá un error similar al siguiente:

```
The bucket you are attempting to access must be addressed using the specified endpoint.
```

De ser posible, asegúrese de que su clúster y el origen de datos estén en la misma región. Puede especificar una región diferente utilizando la opción REGION con el comando COPY.

> **Note**
>
> Si su clúster y el origen de datos están en regiones de AWS diferentes, se le cobrará el costo por la transferencia de datos. También tiene una latencia más alta.

Error del comando COPY

Ejecute la consulta STL_LOAD_ERRORS para descubrir los errores que ocurrieron durante cargas específicas. Para obtener más información, consulte STL_LOAD_ERRORS.

## Tiempo de carga excesivo

La operación de carga puede tardar demasiado por los siguientes motivos. Le sugerimos los siguientes enfoques de resolución de problemas.

Carga de datos de un único archivo mediante el comando COPY

Divida los datos de su carga en archivos múltiples. Cuando carga todos los datos desde un único archivo grande, Amazon Redshift se ve obligado a realizar una carga serializada, que es mucho más lenta. La cantidad de archivos debe ser múltiplo de la cantidad de sectores de su clúster y los archivos deben ser, aproximadamente, del mismo tamaño, entre 1 MB y 1 GB después de la compresión. Para obtener más información, consulte Prácticas recomendadas de Amazon Redshift para el diseño de consultas.

Operación de carga con comandos COPY múltiples

Si utiliza múltiples comandos COPY en simultáneo para cargar una tabla desde varios archivos, Amazon Redshift se ve obligado a realizar una carga serializada, que es mucho más lenta. En ese caso, utilice un único comando COPY.

## Datos de carga incorrectos

La operación COPY puede cargar datos incorrectos de las siguientes maneras. Le sugerimos los siguientes enfoques de resolución de problemas.

Carga de archivos incorrectos

Si se utiliza un prefijo de objeto para especificar archivos de datos, es posible que se lean archivos no deseados. En lugar de ello, utilice un archivo de manifiesto para especificar exactamente qué archivos desea cargar. Para obtener más información, consulte la opción copy_from_s3_manifest_file correspondiente el comando COPY y Example: COPY from Amazon S3 using a manifest en los ejemplos de COPY.

## Configuración del parámetro de tamaño de búsqueda de la JDBC

De manera predeterminada, el controlador Java Database Connectivity (JDBC, Conectividad de base de datos java) recopila todos los resultados de una consulta a la vez. Por este motivo, cuando intenta recuperar un conjunto grande de resultados con una conexión JDBC, es posible que se produzca un error de memoria insuficiente del cliente. Para habilitar su cliente y recuperar conjuntos de resultados en lotes, en lugar de hacerlo en una única búsqueda del tipo "todo o nada", configure el parámetro de tamaño de búsqueda de la JDBC en su aplicación cliente.

> **Note**
>
> El tamaño de búsqueda no es compatible con la Open Database Connectivity (ODBC, Conectividad de base de datos abierta).

Para conseguir el mejor rendimiento, configure el tamaño de búsqueda con el valor máximo que no genere errores de memoria insuficiente. Un valor de tamaño de búsqueda inferior genera más vueltas en el servidor, lo que prolonga los tiempos de ejecución. El servidor reserva recursos, incluido el slot de consultas de Workload Management (WLM, Administración de cargas de trabajo) y la memoria asociada hasta que el cliente recupera todo el conjunto de resultados o hasta que se

cancela la consulta. Cuando ajusta el tamaño de búsqueda como corresponde, esos recursos se liberan con mayor rapidez, lo que hace que queden disponibles para otras consultas.

> **Note**
>
> Si necesita extraer grandes conjuntos de datos, le recomendamos que utilice una instrucción UNLOAD para transferir los datos a Amazon S3. Cuando usa UNLOAD, los nodos de computación trabajan en paralelo para acelerar la transferencia de datos.

Para obtener más información acerca de la configuración de los parámetros de tamaño de búsqueda de JDBC, visite Getting results based on a cursor en la documentación de PostgreSQL.

# Administración de la carga de trabajo

Es posible configurar la WLM de Amazon Redshift para su ejecución con WLM automática o manual.

Con Amazon Redshift, puede administrar y priorizar las consultas simultáneas y las cargas de trabajo de los usuarios para optimizar el rendimiento y el uso de los recursos. La administración de la carga de trabajo (WLM) le permite definir colas, grupos de usuarios y otras estructuras para controlar los recursos asignados a los distintos tipos de consultas o usuarios.

En las siguientes secciones se describen las características específicas de administración de la carga de trabajo de Amazon Redshift y se explica cómo configurarlas y supervisarlas.

WLM automática

Para maximizar el rendimiento del sistema y utilizar los recursos de forma eficaz, puede habilitar a Amazon Redshift para que administre el modo en que se reparten los recursos para ejecutar consultas simultáneas con la WLM automática. La WLM automática administra los recursos necesarios para la ejecución de consultas. Amazon Redshift determina cuántas consultas se ejecutan de forma simultánea y cuánta memoria se asigna a cada consulta enviada. Utilice WLM automática cuando quiera que Amazon Redshift administre el modo en que se dividen los recursos para ejecutar consultas simultáneas. Para obtener más información, consulte Implementación de WLM automática.

Trabajando con el escalado de simultaneidad y la WLM automática, puede permitir un número prácticamente ilimitado de consultas y usuarios simultáneos con un rendimiento de las consultas rápido en todo momento. Para obtener más información, consulte Escalado de simultaneidad.

> **Note**
>
> En la mayoría de los casos le recomendamos que utilice WLM automática. Si utiliza WLM manual y desea migrar a la WLM automática, consulte Migración desde WLM manual a WLM automática.

Con WLM automática, es posible definir las prioridades de consulta para las cargas de trabajo de una cola. Para obtener más información acerca de la prioridad de una consulta, consulte Prioridad de consulta.

WLM manual

Puede tener varias sesiones o usuarios que estén ejecutando consultas al mismo tiempo. Algunas consultas pueden consumir recursos del clúster durante largos periodos y afectar el rendimiento de otras. WLM manual puede ayudar a administrar esto para casos de uso especializados. Utilice WLM manual cuando desee tener más control sobre la simultaneidad.

Puede administrar el rendimiento del sistema si modifica la configuración de la WLM para crear colas separadas para las consultas de ejecución prolongada y de ejecución corta. En el tiempo de ejecución, puede dirigir las consultas a estas colas según los grupos de usuarios o de consultas.

Puede configurar reglas para dirigir consultas a colas particulares en función del usuario que ejecuta la consulta o las etiquetas que especifique. También puede configurar la cantidad de memoria asignada a cada cola, de modo que las consultas grandes se ejecuten en las colas que tienen más memoria que otras. También puede configurar una regla de monitorización de consultas (QMR) para limitar las consultas de ejecución prolongada. Para obtener más información, consulte Implementación de WLM manual.

> **Note**
>
> Le recomendamos que configure las colas de consultas de WLM manual con un total de 15 o menos slots de consulta. Para obtener más información, consulte Nivel de simultaneidad.

Tenga en cuenta que, en lo que respecta a la configuración de WLM manual, el máximo de ranuras que puede asignar a una cola es de 50. No obstante, esto no significa que en una configuración de WLM automática, un clúster de Amazon Redshift ejecute siempre 50 consultas simultáneamente. Esto puede cambiar, según las necesidades de memoria u otros tipos de asignación de recursos en el clúster.

Temas


- Cambio del modo de WLM
- Modificación de la configuración de WLM
- Implementación de WLM automática
- Implementación de WLM manual
- Escalado de simultaneidad
- Aceleración de consultas cortas
- Reglas de asignación de colas de WLM
- Asignación de consultas a las colas

- Propiedades de configuración dinámicas y estáticas de WLM
- Reglas de monitoreo de consultas de WLM
- Tablas y vistas de sistema de WLM

# Cambio del modo de WLM

Puede habilitar WLM manual o automática mediante la consola de Amazon Redshift:

1. Elija Switch WLM mode (Cambiar modo de WLM).
2. Para establecer a WLM automática, elija WLM automática. Con esta opción, se utilizan hasta ocho colas para administrar consultas y los campos Memory (Memoria) y Concurrency on main (Simultaneidad en principal) se establecen ambos en Auto (Automático). Además, la prioridad predeterminada de las consultas se establece en Normal.
3. Para habilitar la configuración manual mediante la consola de Amazon Redshift, cambie a WLM manual. Con esta opción, especifica las colas que se utilizan para administrar consultas y los valores de los campos Memory (Memoria) y Concurrency on main (Simultaneidad en principal). Con una configuración manual, puede configurar hasta ocho colas de consultas y definir el número de consultas que se pueden ejecutar simultáneamente en cada una de estas colas.

# Modificación de la configuración de WLM

La manera más fácil de modificar la configuración de WLM es a través de la consola de Amazon Redshift. También puede utilizar la AWS CLI o la API de Amazon Redshift.

Al cambiar el clúster entre WLM automática y manual, el clúster se pone en estado `pending reboot`. El cambio no surte efecto hasta que se reinicia el siguiente clúster.

Para obtener información detallada sobre cómo modificar las configuraciones de WLM, consulte Configuración de la administración de cargas de trabajo en la Guía de administración de Amazon Redshift.

## Migración desde WLM manual a WLM automática

Para maximizar el rendimiento del sistema y utilizar recursos de manera más eficaz, le recomendamos que configure WLM automáticas para sus colas. Plantéese adoptar el enfoque siguiente para configurar una transición suave desde WLM manual a WLM automática.

Para migrar de WLM manual a WLM automática y utilizar prioridades de consulta, le recomendamos crear un nuevo grupo de parámetros y, a continuación, asociarlo a su clúster. Para obtener más información, consulte Grupos de parámetros de Amazon Redshift en la Guía de administración de Amazon Redshift.

> **Important**
>
> Para cambiar el grupo de parámetros o para cambiar de WLM manual a automática se requiere un reinicio del clúster. Para obtener más información, consulte Propiedades de configuración dinámicas y estáticas de WLM.

Tomemos un ejemplo donde hay tres colas de WLM manual. Una para una carga de trabajo de ETL, una carga de trabajo de análisis y una carga de trabajo de ciencia de datos. La carga de trabajo de ETL se ejecuta cada 6 horas, la carga de trabajo de análisis se ejecuta durante el día y la carga de trabajo de ciencia de datos puede tener un pico en cualquier momento. Con la WLM manual, especifica la memoria y la simultaneidad que obtiene cada cola de carga de trabajo en función de su comprensión de la importancia de cada carga de trabajo para el negocio. Especificar la memoria y la simultaneidad no solo es difícil de averiguar, sino que también da lugar a que los recursos del clúster se particionen estáticamente y, por tanto, se desaprovechen cuando solo está funcionando un subconjunto de las cargas de trabajo.

Puede utilizar WLM automático con prioridades de consulta para indicar las prioridades relativas de las cargas de trabajo, evitando los problemas anteriores. Para este ejemplo, siga estos pasos:

- Cree un nuevo grupo de parámetros y cambie al modo Auto WLM (WLM automática).
- Añada colas para cada una de las tres cargas de trabajo: carga de trabajo de ETL, carga de trabajo de análisis y carga de trabajo de ciencia de datos. Utilice los mismos grupos de usuarios en cada carga de trabajo que usó en el modo Manual WLM (WLM manual).
- Establezca la prioridad de la carga de trabajo de ETL en `High`, la carga de trabajo de análisis en `Normal` y la ciencia de datos en `Low`. Estas prioridades reflejan sus prioridades de negocio para las distintas cargas de trabajo o grupos de usuarios.
- También puede habilitar el escalado de simultaneidad para la cola de análisis o de ciencia de datos de modo que las consultas en estas colas consigan un rendimiento coherente incluso cuando la carga de trabajo de ETL se ejecuta cada 6 horas.

Con las prioridades de consulta, cuando solo se ejecuta la carga de trabajo de análisis en el clúster, este obtiene todo el sistema para sí. Así se obtiene un alto rendimiento con una mejor utilización del sistema. Sin embargo, cuando se inicia la carga de trabajo de ETL, obtiene el derecho de paso ya que tiene una prioridad más alta. Las consultas que se ejecutan como parte de la carga de trabajo de ETL obtienen prioridad durante la admisión, además de la asignación de recursos preferente después de que se admitan. En consecuencia, la carga de trabajo de ETL tiene un rendimiento predecible con independencia de los demás elementos que se estén ejecutando en el sistema. El rendimiento predecible para una carga de trabajo de alta prioridad conlleva el costo de otras cargas de trabajo de prioridad inferior que se ejecutan durante más tiempo ya que sus consultas están esperando a que se completen consultas más importantes. O, dado que están obteniendo una fracción de recursos inferior cuando se ejecutan de forma simultánea con consultas de prioridad más alta. Los algoritmos de programación utilizados por Amazon Redshift facilitan que las consultas de prioridad más baja no sufran de agotamiento, sino que sigan avanzando, aunque a un ritmo más lento.

> **Note**
>
> - El campo de tiempo de espera no está disponible en WLM automática. En su lugar, utilice la regla de QMR, `query_execution_time`. Para obtener más información, consulte Reglas de monitoreo de consultas de WLM.
> - La acción de QMR, HOP, no es aplicable a la WLM automática. En su lugar, utilice la acción `change priority`. Para obtener más información, consulte Reglas de monitoreo de consultas de WLM.
> - Los clústeres utilizan las colas de WLM automáticas y manuales de forma diferente, lo que puede crear confusión con sus configuraciones. Por ejemplo, puede configurar la propiedad de prioridad en las colas de WLM automáticas, pero no en las colas de WLM manuales. Por ello, evite mezclar colas de WLM automáticas y colas de WLM manuales en un grupo de parámetros. En su lugar, cree un grupo de parámetros al migrar a la WLM automática.

## Implementación de WLM automática

Con la administración de cargas de trabajo (WLM) automática, Amazon Redshift administra la simultaneidad de consultas y la asignación de memoria. Puede crear hasta ocho colas con los

identificadores de clase de servicio 100-107. Cada cola tiene una prioridad. Para obtener más información, consulte Prioridad de consulta.

La WLM automática determina la cantidad de recursos que necesitan las consultas y ajusta la simultaneidad en función de la carga de trabajo. Cuando en el sistema hay consultas que requieren grandes cantidades de recursos (por ejemplo, combinaciones hash entre tablas grandes), la simultaneidad es inferior. Cuando se envían consultas más ligeras (como por ejemplo inserciones, eliminaciones, exámenes o agregaciones sencillas), la simultaneidad es superior.

La WLM automática es independiente de la aceleración de consultas cortas (SQA) y evalúa las consultas de forma diferente. La WLM automática y la SQA funcionan conjuntamente para permitir que las consultas ligeras y de corta ejecución se completen incluso si hay activas consultas de uso intensivo de recursos y de ejecución prolongada. Para obtener más información acerca de SQA, consulte Aceleración de consultas cortas.

Amazon Redshift habilita la WLM automática a través de grupos de parámetros:

- Si los clústeres utilizan el grupo de parámetros predeterminado, Amazon Redshift habilita la WLM automática para ellos.
- Si los clústeres utilizan grupos de parámetros personalizados, puede configurar los clústeres para habilitar la WLM automática. Le recomendamos que cree un grupo de parámetros independiente para la configuración de WLM automática.

Para configurar WLM, edite el parámetro `wlm_json_configuration` en un grupo de parámetros que se puede asociar a uno o varios clústeres. Para obtener más información, consulte Modificación de la configuración de WLM.

Puede definir colas de consultas dentro de la configuración de WLM. Puede añadir colas de consultas adicionales a la configuración de WLM predeterminada, hasta llegar a un total de ocho colas de usuario. Para cada cola de consultas, puede configurar lo siguiente:

- Priority (Prioridad)
- Modo de escalado de simultaneidad
- Grupos de usuarios
- Grupos de consultas
- Reglas de monitorización de consultas

## Priority (Prioridad)

Puede definir la importancia relativa de las consultas en una carga de trabajo estableciendo un valor de prioridad. La prioridad se especifica para una cola y la heredan todas las consultas asociadas a la cola. Para obtener más información, consulte Prioridad de consulta.

## Modo de escalado de simultaneidad

Cuando el escalado de simultaneidad está habilitado, Amazon Redshift agrega capacidad del clúster de manera automática si se necesita para procesar un aumento de las consultas de lectura y escritura simultáneas. Los usuarios ven siempre los datos más actualizados, tanto si las consultas se ejecutan en el clúster principal como si se ejecutan en un clúster de escalado de simultaneidad.

Puede administrar qué consultas se envían al clúster de escalado de simultaneidad configurando colas de WLM. Cuando habilita el escalado de simultaneidad para una cola, las consultas elegibles se envían al clúster de escalado de simultaneidad en lugar de esperar en la cola. Para obtener más información, consulte Escalado de simultaneidad.

## Grupos de usuarios

Puede asignar un conjunto de grupos de usuarios a una cola especificando cada nombre de grupo de usuarios o utilizando comodines. Cuando un miembro de un grupo de usuarios de la lista ejecuta una consulta, esa consulta se ejecuta en la cola correspondiente. No hay un límite establecido en cuanto al número de grupos de usuarios que se pueden asignar a una cola. Para obtener más información, consulte Asignación de consultas a las colas en función de los grupos de usuarios.

## Roles de usuario

Puede asignar un conjunto de roles de usuarios a una cola especificando cada nombre de rol de usuarios o utilizando comodines. Cuando un miembro de un rol de usuario de la lista ejecuta una consulta, esa consulta se ejecuta en la cola correspondiente. No hay un límite establecido en cuanto al número de roles de usuarios que se pueden asignar a una cola. Para obtener más información, consulte Asignación de consultas a las colas en función de los roles de usuarios.

## Grupos de consultas

Puede asignar un conjunto de grupos de consultas a una cola especificando cada nombre de grupo de consultas o utilizando comodines. Un grupo de consultas es sencillamente una etiqueta. En tiempo de ejecución, puede asignar la etiqueta de grupo de consultas a un conjunto de

consultas. Cualquier consulta que se asigne a un grupo de consultas de la lista se ejecuta en la cola correspondiente. No hay un límite establecido para el número de grupos de consultas que se pueden asignar a una cola. Para obtener más información, consulte Asignación de una consulta a un grupo de consultas.

## Caracteres comodín

Si los comodines están habilitados en la configuración de la cola de WLM, puede asignar grupos de usuarios y de consultas a una cola individualmente o utilizando los comodines de estilo shell Unix. La coincidencia de patrones no distingue entre mayúsculas y minúsculas.

Por ejemplo, el comodín «*» coincide con cualquier número de caracteres. Así, si añade `dba_*` a la lista de grupos de usuario para una cola, cualquier consulta ejecutada por un usuario que pertenece a un grupo con un nombre que comienza por `dba_` se asigna a esa cola. Algunos ejemplos son `dba_admin` o `DBA_primary`. El comodín «?» coincide con cualquier carácter individual. Así, si la cola incluye el grupo de usuarios `dba?1`, entonces los grupos de usuarios denominados `dba11` y `dba21` coinciden, pero `dba12` no coincide.

De forma predeterminada, los comodines no están habilitados.

## Reglas de monitorización de consultas

Las reglas de monitorización de consultas definen los límites de rendimiento basados en métricas para las colas de WLM y especifican la acción que se debe realizar cuando una consulta va más allá de esos límites. Por ejemplo, para una cola dedicada a consultas de ejecución corta, puede crear una regla que cancele las consultas que se ejecuten durante más de 60 segundos. Para hacer un seguimiento de las consultas mal diseñadas, puede disponer de otra regla que registre las consultas que contienen bucles anidados. Para obtener más información, consulte Reglas de monitoreo de consultas de WLM.

## Comprobación de WLM automática

Para saber si la WLM automática está habilitada, ejecute la consulta siguiente. Si la consulta devuelve al menos una fila, la WLM automática está habilitada.

```
select * from stv_wlm_service_class_config
where service_class >= 100;
```

En la consulta siguiente, se muestra el número de consultas que pasaron por cada cola de consultas (clase de servicio). También se muestra el tiempo de ejecución medio, el número de consultas con

un tiempo de espera establecido en el percentil 90 y el tiempo de espera medio. Las consultas de WLM automática utilizan las clases de servicio 100 a 107.

```
select final_state, service_class, count(*), avg(total_exec_time),
percentile_cont(0.9) within group (order by total_queue_time), avg(total_queue_time)
from stl_wlm_query where userid >= 100 group by 1,2 order by 2,1;
```

Para saber qué consultas se han ejecutado mediante la WLM automática y se han completado correctamente, ejecute la consulta siguiente.

```
select a.queue_start_time, a.total_exec_time, label, trim(querytxt)
from stl_wlm_query a, stl_query b
where a.query = b.query and a.service_class >= 100 and a.final_state = 'Completed'
order by b.query desc limit 5;
```

## Prioridad de consulta

Con Amazon Redshift, puede administrar la priorización de consultas y la asignación de recursos entre consultas y cargas de trabajo simultáneas mediante la administración de la carga de trabajo (WM). En las siguientes secciones se detalla cómo configurar las colas de consultas de WM, definir las propiedades de las colas, como la asignación de memoria y el escalado de simultaneidad, e implementar reglas de prioridad adaptadas a sus requisitos de carga de trabajo.

No todas las consultas tienen la misma importancia y, con frecuencia, el rendimiento de una carga de trabajo o de un conjunto de usuarios podría ser más importante. Si ha habilitado la WLM automática, puede definir la importancia relativa de las consultas en una carga de trabajo estableciendo un valor de prioridad. La prioridad se especifica para una cola y la heredan todas las consultas asociadas a la cola. Asocie las consultas a una cola asignando grupos de usuarios y grupos de consultas a la cola. Puede establecer las propiedades siguientes (enumeradas de prioridad más alta a más baja):

1. `HIGHEST`
2. `HIGH`
3. `NORMAL`
4. `LOW`
5. `LOWEST`

Los administradores utilizan estas prioridades para mostrar la importancia relativa de las cargas de trabajo cuando hay consultas con distintas prioridades que compiten por los mismos

recursos. Amazon Redshift utiliza la prioridad cuando permite que se ingresen consultas en el sistema y también para determinar la cantidad de recursos asignados a una consulta. De forma predeterminada, las consultas se ejecutan con su prioridad establecida en `NORMAL`.

Una prioridad adicional, `CRITICAL`, que es una prioridad mayor que `HIGHEST`, está disponible para los superusuarios. Para establecer esta prioridad, puede utilizar las funciones CHANGE_QUERY_PRIORITY, CHANGE_SESSION_PRIORITY y CHANGE_USER_PRIORITY. Para conceder a una base de datos permisos de usuario para utilizar estas funciones, puede crear un procedimiento almacenado y conceder un permiso a un usuario. Para ver un ejemplo, consulta CHANGE_SESSION_PRIORITY.

> **Note**
>
> Solo se puede ejecutar una consulta `CRITICAL` a la vez.
> Las reversiones siempre tienen prioridad CRÍTICA.

Veamos un ejemplo donde la prioridad de una carga de trabajo de extracción, transformación, carga (ETL) es superior a la prioridad de la carga de trabajo de análisis. La carga de trabajo ETL se ejecuta cada seis horas y la carga de trabajo de análisis se ejecuta a lo largo del día. Cuando solo se ejecuta la carga de trabajo de análisis en el clúster, consigue todo el sistema para sí produciendo un rendimiento alto con una utilización óptima del sistema. Sin embargo, cuando se inicia la carga de trabajo de ETL, obtiene el derecho de paso ya que tiene una prioridad más alta. Las consultas que se ejecutan como parte de la carga de trabajo de ETL obtienen el derecho de paso durante la admisión además de la asignación de recursos preferente después de que se admitan. En consecuencia, la carga de trabajo de ETL tiene un rendimiento predecible con independencia de los demás elementos que se estén ejecutando en el sistema. De este modo, proporciona un rendimiento predecible y la capacidad para los administradores de proporcionar acuerdos de nivel de servicio (SLA) para sus usuarios empresariales.

Dentro de un clúster determinado, el rendimiento predecible para una carga de trabajo de alta prioridad se hace a expensas de otras cargas de trabajo de menor prioridad. Las cargas de trabajo de prioridad inferior podrían tardar más tiempo en ejecutarse ya que sus consultas esperan a que se completen consultas más importantes. O, podrían tardar más tiempo, dado que están obteniendo una fracción de recursos inferior cuando se ejecutan de forma simultánea con consultas de prioridad más alta. Las consultas de prioridad inferior no sufren el agotamiento, sino que siguen avanzando a un ritmo más lento.

En el ejemplo anterior, el administrador puede habilitar el escalado simultáneo para la carga de trabajo de análisis. Esto permite que la carga de trabajo mantenga su rendimiento, incluso aunque la carga de trabajo de ETL se está ejecutando con prioridad alta.

## Configuración de prioridad de colas

Si ha habilitado la WLM automática, cada cola tiene un valor de prioridad. Las consultas se direccionan a las colas en función de los grupos de usuarios y de consultas. Comience con una prioridad de cola configurada como `NORMAL`. Defina la prioridad más alta o más baja en función de la carga de trabajo asociada a los grupos de usuario de la cola y los grupos de consulta.

Puede cambiar la prioridad de una cola en la consola de Amazon Redshift. En la consola de Amazon Redshift, la página Workload Management (Administración de cargas de trabajo) muestra las colas y habilita la edición de sus propiedades, como Priority (Prioridad). Para establecer la prioridad utilizando las operaciones de la CLI o de la API, utilice el parámetro `wlm_json_configuration`. Para obtener más información, consulte Configuración de la administración de cargas de trabajo en la Guía de administración de Amazon Redshift.

El siguiente ejemplo de `wlm_json_configuration` define tres grupos de usuarios (`ingest`, `reporting` y `analytics`). Las consultas enviadas de los usuarios de uno de estos grupos se ejecutan con prioridad `highest`, `normal` y `low` respectivamente.

```
[
    {
        "user_group": [
            "ingest"
        ],
        "priority": "highest",
        "queue_type": "auto"
    },
    {
        "user_group": [
            "reporting"
        ],
        "priority": "normal",
        "queue_type": "auto"
    },
    {
        "user_group": [
            "analytics"
        ],
```

```
        "priority": "low",
        "queue_type": "auto",
        "auto_wlm": true
    }
]
```

## Cambio de la prioridad de consulta con reglas de monitoreo de consultas

Las reglas de monitorización de consultas (QMR) le permiten cambiar la prioridad de una consulta en función de su comportamiento mientras está en ejecución. Esto se hace especificando el atributo de prioridad en un predicador de QMR además de una acción. Para obtener más información, consulte Reglas de monitoreo de consultas de WLM.

Por ejemplo, puede definir una regla para cancelar cualquier consulta clasificada como prioridad `high` que se ejecute durante más de 10 minutos.

```
"rules" :[
  {
    "rule_name":"rule_abort",
    "predicate":[
      {
        "metric_name":"query_cpu_time",
        "operator":">",
        "value":600
      },
      {
        "metric_name":"query_priority",
        "operator":"=",
        "value":"high"
      }
    ],
    "action":"abort"
  }
]
```

Otro ejemplo consiste en definir una regla para cambiar la prioridad de consulta a `lowest` para cualquier consulta con la prioridad actual `normal` que vierte más de 1 TB en disco.

```
"rules":[
  {
    "rule_name":"rule_change_priority",
```

```
    "predicate":[
      {
        "metric_name":"query_temp_blocks_to_disk",
        "operator":">",
        "value":1000000
      },
      {
        "metric_name":"query_priority",
        "operator":"=",
        "value":"normal"
      }
    ],
    "action":"change_query_priority",
    "value":"lowest"
  }
]
```

## Monitoreo de la prioridad de consulta

Para mostrar la prioridad para consultas en espera y en ejecución, vea la columna `query_priority` en la tabla del sistema stv_wlm_query_state.

```
query    | service_cl | wlm_start_time             | state            | queue_time |
 query_priority
---------+------------+----------------------------+------------------+------------
+----------------
2673299  | 102        | 2019-06-24 17:35:38.866356 | QueuedWaiting    | 265116     |
 Highest
2673236  | 101        | 2019-06-24 17:35:33.313854 | Running          | 0          |
 Highest
2673265  | 102        | 2019-06-24 17:35:33.523332 | Running          | 0          |
 High
2673284  | 102        | 2019-06-24 17:35:38.477366 | Running          | 0          |
 Highest
2673288  | 102        | 2019-06-24 17:35:38.621819 | Running          | 0          |
 Highest
2673310  | 103        | 2019-06-24 17:35:39.068513 | QueuedWaiting    | 62970      |
 High
2673303  | 102        | 2019-06-24 17:35:38.968921 | QueuedWaiting    | 162560     |
 Normal
2673306  | 104        | 2019-06-24 17:35:39.002733 | QueuedWaiting    | 128691     |
 Lowest
```

Para enumerar la prioridad de consulta para consultas completadas, vea la columna `query_priority` en la tabla del sistema stl_wlm_query.

```
select query, service_class as svclass, service_class_start_time as starttime,
 query_priority
from stl_wlm_query order by 3 desc limit 10;
```

```
  query   | svclass |         starttime          |    query_priority
---------+---------+----------------------------+----------------------
 2723254 |     100 | 2019-06-24 18:14:50.780094 | Normal
 2723251 |     102 | 2019-06-24 18:14:50.749961 | Highest
 2723246 |     102 | 2019-06-24 18:14:50.725275 | Highest
 2723244 |     103 | 2019-06-24 18:14:50.719241 | High
 2723243 |     101 | 2019-06-24 18:14:50.699325 | Low
 2723242 |     102 | 2019-06-24 18:14:50.692573 | Highest
 2723239 |     101 | 2019-06-24 18:14:50.668535 | Low
 2723237 |     102 | 2019-06-24 18:14:50.661918 | Highest
 2723236 |     102 | 2019-06-24 18:14:50.643636 | Highest
```

Para optimizar el rendimiento de la carga de trabajo, es posible que Amazon Redshift modifique la prioridad de las consultas enviadas por el usuario. Amazon Redshift utiliza algoritmos avanzados de machine learning para determinar cuáles son los momentos en que esta optimización beneficia a la carga de trabajo y la aplica automáticamente cuando se cumplen todas las condiciones que se indican a continuación.

- La WLM automática está habilitada.
- Solo se define una cola de WLM.
- No ha definido reglas de monitoreo de consultas (QMR) que establezcan la prioridad de las consultas. Estas reglas incluyen la métrica de QMR `query_priority` o la acción de QMR `change_query_priority`. Para obtener más información, consulte Reglas de monitoreo de consultas de WLM.

## Implementación de WLM manual

Con la WLM manual, puede administrar el rendimiento del sistema y la experiencia de los usuarios mediante la modificación de la configuración de la WLM para crear colas independientes para las consultas de ejecución prolongada y las de ejecución corta.

Cuando los usuarios ejecutan consultas en Amazon Redshift, estas se dirigen a las colas de consultas. Cada cola de consultas contiene un número de slots de consultas. A cada cola se le asigna una parte de la memoria disponible del clúster. La memoria de una cola se divide entre los slots de consultas de la cola. Puede habilitar Amazon Redshift para administrar la simultaneidad de las consultas con la WLM automática. Para obtener más información, consulte Implementación de WLM automática.

O puede configurar las propiedades de WLM para cada cola de consulta. Lo hace para especificar la forma en que se asigna la memoria entre ranuras y cómo se pueden dirigir las consultas a colas específicas en tiempo de ejecución. También es posible configurar propiedades de WLM para cancelar consultas de ejecución prolongada.

De forma predeterminada, Amazon Redshift configura las siguientes colas de consultas:

- Una cola de superusuario

  La cola de superusuario está reservada únicamente para superusuarios y no se puede configurar. Utilice esta cola solo cuando necesite ejecutar consultas que afectan al sistema o para solucionar problemas. Por ejemplo, utilice esta cola cuando necesite cancelar una consulta de ejecución prolongada de un usuario o para añadir usuarios a la base de datos. No la utilice para realizar consultas de rutina. La cola no aparece en la consola, pero sí lo hace en las tablas de sistema de la base de datos como la quinta cola. Para ejecutar una consulta en la cola de superusuario, es necesario que un usuario inicie sesión como superusuario y debe ejecutar la consulta mediante el grupo de consultas predefinido `superuser`.
- Una cola de usuario predeterminada

  La cola predeterminada está configurada inicialmente para ejecutar cinco consultas de forma simultánea. Al usar WLM manual, puede cambiar las propiedades de simultaneidad, tiempo de espera y asignación de memoria para la cola predeterminada, pero no puede especificar grupos de usuarios ni de consultas. La cola predeterminada debe ser la última cola en la configuración de WLM. Cualquier consulta que no se direccione a otras colas se ejecuta en la cola predeterminada.

Las colas de consultas se definen en la configuración de WLM. La configuración de WLM es un parámetro modificable (`wlm_json_configuration`) de un grupo de parámetros, que puede estar asociado a uno o más clústeres. Para obtener más información, consulte Configuración de la administración de cargas de trabajo en la Guía de administración de Amazon Redshift.

Puede añadir colas de consultas adicionales a la configuración de WLM predeterminada, hasta llegar a un total de ocho colas de usuario. Para cada cola de consultas, puede configurar lo siguiente:

- Modo de escalado de simultaneidad
- Nivel de simultaneidad
- Grupos de usuarios
- Grupos de consultas
- Porcentaje de memoria de WLM por utilizar
- Tiempo de espera de WLM
- Salto de cola de consultas de WLM
- Reglas de monitorización de consultas

## Modo de escalado de simultaneidad

Cuando el escalado de simultaneidad está habilitado, Amazon Redshift agrega capacidad del clúster de manera automática si se necesita para procesar un aumento de las consultas de lectura y escritura simultáneas. Los usuarios ven siempre los datos más actualizados, tanto si las consultas se ejecutan en el clúster principal como si se ejecutan en un clúster de escalado de simultaneidad.

Puede administrar qué consultas se envían al clúster de escalado de simultaneidad configurando colas de WLM. Cuando habilita el escalado de simultaneidad para una cola, las consultas elegibles se envían al clúster de escalado de simultaneidad en lugar de esperar en la cola. Para obtener más información, consulte Escalado de simultaneidad.

## Nivel de simultaneidad

Las consultas de una cola se ejecutan simultáneamente hasta que se alcanza el número de slots de consultas de WLM o el nivel de simultaneidad definido para esa cola. Las consultas posteriores esperan en la cola.

> **Note**
>
> El nivel de simultaneidad de WLM es distinto del número de conexiones de usuario simultáneas que se pueden realizar en un clúster. Para obtener más información, consulte Conexión a un clúster en la Guía de administración de Amazon Redshift.

En una configuración de WLM automática, que es la recomendada, el nivel de simultaneidad se establece a Automático. Amazon Redshift asigna memoria de forma dinámica a las consultas y, posteriormente, determina cuántas se ejecutarán de forma simultánea. Se basa en los recursos necesarios tanto para las consultas en ejecución como para las que están en cola. No se puede configurar WLM automática. Para obtener más información, consulte Implementación de WLM automática.

En una configuración de WLM manual, Amazon Redshift asigna de forma estática una cantidad fija de memoria a cada cola. La memoria de la cola se divide en partes iguales entre las ranuras de consulta. A modo de ejemplo, si a una cola se le asigna el 20 % de la memoria de un clúster y tiene 10 ranuras, a cada consulta se le asigna el 2 % de la memoria del clúster. La asignación de memoria permanece fija independientemente del número de consultas que se ejecuten de forma simultánea. Debido a esta asignación de memoria fija, las consultas que se ejecutan completamente en la memoria cuando el número de ranuras es 5, puede que deba escribir resultados intermedios en el disco si el número de ranuras aumenta a 20. En este caso, la cuota de memoria de la cola correspondiente a cada consulta se reduce de un 1/5 a un 1/20. La E/S adicional del disco puede degradar el rendimiento.

El número máximo de ranuras de consultas para todas las colas definidas por el usuario es de 50. Esto limita el total de ranuras para todas las colas, incluida la cola predeterminada. La única cola que no está sujeta al límite es la cola de superusuario reservada.

De forma predeterminada, las colas de WLM manuales tienen un nivel de simultaneidad de 5. La carga de trabajo puede beneficiarse de un nivel de simultaneidad más alto en determinados casos, como los siguientes:

- Si hay muchas consultas pequeñas que tienen que esperar a que se ejecuten las consultas largas, cree una cola independiente con un mayor número de slots y asigne las consultas pequeñas a esa cola. Una cola con un nivel de simultaneidad más alto tiene menos memoria asignada a cada slot de consultas, pero las consultas más pequeñas requieren menos memoria.

> **Note**
>
> Si habilita la aceleración de consultas cortas (SQA), WLM dará prioridad automáticamente a las consultas cortas frente a las largas, por lo que no será necesario que separe la cola de las consultas cortas en la mayoría de los flujos de trabajo. Para obtener más información, consulte Aceleración de consultas cortas.

- Si varias colas tienen acceso a los datos de un mismo sector, configure una cola de WLM independiente para ejecutar esas consultas de manera simultánea. Amazon Redshift asigna consultas simultáneas a diferentes sectores, lo que permite que varias consultas se ejecuten en paralelo en distintos sectores. Por ejemplo, si una consulta es un conjunto simple con un predicado en la clave de distribución, los datos para la consulta se ubican en un sector único.

## Un ejemplo de WLM manual

Este ejemplo es un escenario de WLM sencillo y manual para mostrar cómo se pueden asignar las ranuras y la memoria. La WLM manual se implementa con tres colas, que son las siguientes:

- cola de ingesta de datos: está configurada para ingerir datos. Se le asigna el 20 % de la memoria del clúster y tiene 5 ranuras. Posteriormente, se pueden ejecutar 5 consultas de forma simultánea en la cola y a cada una se le asigna el 4 % de la memoria.
- cola de científicos de datos: está diseñada para consultas que consumen mucha memoria. Se le asigna el 40 % de la memoria del clúster y tiene 5 ranuras. Posteriormente, se pueden ejecutar 5 consultas de forma simultánea y a cada una se le asigna el 8 % de la memoria.
- cola predeterminada: está diseñada para la mayoría de los usuarios de la organización. Se incluyen los grupos de ventas y contabilidad que normalmente tienen consultas de corta o mediana duración que no son complicadas. Se le asigna el 40 % de la memoria del clúster y tiene 40 ranuras. Se pueden ejecutar 40 consultas simultáneamente en esta cola y a cada consulta se le asigna el 1 % de la memoria. Es el número máximo de ranuras que se pueden asignar a esta cola, ya que el límite entre todas las colas es de 50.

Si ejecuta WLM automática y su carga de trabajo requiere que más de 15 consultas se ejecuten en paralelo, le recomendamos que active el escalado de simultaneidad. El motivo es que, al aumentar el número de ranuras de consulta por encima de 15, se podría provocar la contención de los recursos del sistema y que se limitara el rendimiento general de un único clúster. Con el escalado de simultaneidad, puede ejecutar cientos de consultas en paralelo hasta alcanzar el número de clústeres de escalado de simultaneidad establecido. max_concurrency_scaling_clusters controla el número de clústeres de escalado de simultaneidad. Para obtener más información acerca del escalado de simultaneidad, consulte Escalado de simultaneidad.

Para obtener más información, consulte Mejora del rendimiento de las consultas.

## Grupos de usuarios

Puede asignar un conjunto de grupos de usuarios a una cola especificando cada nombre de grupo de usuarios o utilizando comodines. Cuando un miembro de un grupo de usuarios de la lista ejecuta una consulta, esa consulta se ejecuta en la cola correspondiente. No hay un límite establecido en cuanto al número de grupos de usuarios que se pueden asignar a una cola. Para obtener más información, consulte Asignación de consultas a las colas en función de los grupos de usuarios.

## Roles de usuario

Puede asignar un conjunto de roles de usuarios a una cola especificando cada nombre de rol de usuarios o utilizando comodines. Cuando un miembro de un rol de usuario de la lista ejecuta una consulta, esa consulta se ejecuta en la cola correspondiente. No hay un límite establecido en cuanto al número de roles de usuarios que se pueden asignar a una cola. Para obtener más información, consulte Asignación de consultas a las colas en función de los roles de usuarios.

## Grupos de consultas

Puede asignar un conjunto de grupos de consultas a una cola especificando cada nombre de grupo de consultas o utilizando comodines. Un grupo de consultas es sencillamente una etiqueta. En tiempo de ejecución, puede asignar la etiqueta de grupo de consultas a un conjunto de consultas. Cualquier consulta que se asigne a un grupo de consultas de la lista se ejecuta en la cola correspondiente. No hay un límite establecido para el número de grupos de consultas que se pueden asignar a una cola. Para obtener más información, consulte Asignación de una consulta a un grupo de consultas.

## Caracteres comodín

Si los comodines están habilitados en la configuración de la cola de WLM, puede asignar grupos de usuarios y de consultas a una cola individualmente o utilizando los comodines de estilo shell Unix. La coincidencia de patrones no distingue entre mayúsculas y minúsculas.

Por ejemplo, el comodín «*» coincide con cualquier número de caracteres. Así, si añade `dba_*` a la lista de grupos de usuario para una cola, cualquier consulta ejecutada por un usuario que pertenece a un grupo con un nombre que comienza por `dba_` se asigna a esa cola. Algunos ejemplos son `dba_admin` o `DBA_primary`. El comodín «?» coincide con cualquier carácter individual. Así, si la cola incluye el grupo de usuarios `dba?1`, entonces los grupos de usuarios denominados `dba11` y `dba21` coinciden, pero `dba12` no coincide.

Los comodines se encuentran desactivados de forma predeterminada.

## Porcentaje de memoria de WLM por utilizar

En una configuración de WLM automática, el porcentaje de memoria se establece en **`auto`**. Para obtener más información, consulte Implementación de WLM automática.

En una configuración de WLM manual, para especificar la cantidad de memoria disponible que se asigna a una consulta, puede establecer el parámetro `WLM Memory Percent to Use`. De forma predeterminada, a cada cola definida por el usuario se le asigna una parte igual de la memoria que esté disponible para consultas definidas por el usuario. Por ejemplo, si tiene cuatro colas definidas por el usuario, a cada cola se le asigna un 25 por ciento de la memoria disponible. La cola de superusuario tiene su propia memoria asignada y no puede modificarse. Para cambiar la asignación, asigne un porcentaje entero de memoria a cada cola, hasta llegar a un total de 100 por ciento. Amazon Redshift administra la memoria que ha quedado sin asignar, la cual se puede asignar de forma temporal a una cola si esta solicita memoria adicional para el procesamiento.

Por ejemplo, si configura cuatro colas, puede asignar la memoria de la siguiente forma: 20 por ciento, 30 por ciento, 15 por ciento, 15 por ciento. El 20 por ciento restante queda sin asignar y lo administra el servicio.

## Tiempo de espera de WLM

El tiempo de espera de WLM (`max_execution_time`) no está disponible. En su lugar, cree una regla de monitorización de consultas (QMR) utilizando `query_execution_time` para limitar el tiempo de ejecución transcurrido de una consulta. Para obtener más información, consulte Reglas de monitoreo de consultas de WLM.

Para limitar la cantidad de tiempo que las consultas de una determinada cola de WLM tienen permitido utilizar, puede establecer el valor de tiempo de espera de WLM para cada cola.
El parámetro de tiempo de espera especifica la cantidad de tiempo, en milisegundos, que Amazon Redshift espera para que se ejecute una consulta antes de cancelarla o saltarla. El tiempo de espera se basa en el tiempo de ejecución de consultas y no incluye el tiempo de espera empleado en una cola.

WLM intenta transferir las instrucciones CREATE TABLE AS (CTAS) y las consultas de solo lectura, como las instrucciones SELECT. Las consultas que no se pueden transferir se cancelan. Para obtener más información, consulte Salto de cola de consultas de WLM.

El tiempo de espera de WLM no aplica a una consulta que haya alcanzado el estado de devolución. Para ver el estado de una consulta, consulte la tabla de sistema STV_WLM_QUERY_STATE. Las instrucciones COPY y las operaciones de mantenimiento, como ALTER, ANALYZE y VACUUM, no están sujetas al tiempo de espera de WLM.

La función de tiempo de espera de WLM es similar al parámetro de configuración statement_timeout. La diferencia es que, en los casos que el parámetro de configuración `statement_timeout` aplica a todo el clúster, el tiempo de espera de WLM es específico de una cola individual en la configuración de WLM.

Si también se especifica statement_timeout, se utiliza el statement_timeout y el tiempo de espera de WLM (max_execution_time) inferiores.

## Reglas de monitorización de consultas

Las reglas de monitorización de consultas definen los límites de rendimiento basados en métricas para las colas de WLM y especifican la acción que se debe realizar cuando una consulta va más allá de esos límites. Por ejemplo, para una cola dedicada a consultas de ejecución corta, puede crear una regla que cancele las consultas que se ejecuten durante más de 60 segundos. Para hacer un seguimiento de las consultas mal diseñadas, puede disponer de otra regla que registre las consultas que contienen bucles anidados. Para obtener más información, consulte Reglas de monitoreo de consultas de WLM.

## Salto de cola de consultas de WLM

Con Amazon Redshift, puede administrar la simultaneidad de la carga de trabajo y la asignación de recursos al habilitar el salto de cola de consultas de WLM (Administración de la carga de trabajo). Esta característica permite que las consultas “salten” temporalmente de una cola asignada a una cola de mayor prioridad cuando hay recursos disponibles, lo que mejora el rendimiento general de las consultas y la utilización del sistema. En las siguientes secciones se proporcionan instrucciones detalladas sobre la configuración y el uso de los saltos de cola de consultas de WLM en Amazon Redshift.

Una consulta puede saltarse porque se ha agotado el tiempo de espera de WLM o por una acción de salto de una regla de monitoreo de consultas (QMR). Solo puede transferir consultas en una configuración de WLM manual.

Cuando se salta una consulta, WLM intenta direccionarla a la siguiente cola coincidente en función de las reglas de asignación de colas de WLM. Si la consulta no coincide con ninguna otra definición de cola, se cancelará. No se asigna a la cola predeterminada.

## Acciones de tiempo de espera de WLM

En la siguiente tabla, se resume el comportamiento de diferentes tipos de consulta con un tiempo de espera de WLM.

| Tipo de consulta | Acción |
| --- | --- |
| INSERT, UPDATE y DELETE | Cancelación |
| Funciones definidas por el usuario (UDF) | Cancelación |
| UNLOAD | Cancelación |
| COPY | Siguen ejecutándose |
| Operaciones de mantenimiento | Siguen ejecutándose |
| Consultas de solo lectura con el estado `returning` | Siguen ejecutándose |
| Consultas de solo lectura con el estado `running` | Se reasignan o reinicia |
| CREATE TABLE AS (CTAS), SELECT INTO | Se reasignan o reinicia |

## Salto de cola por un tiempo de espera de WLM

WLM transfiere los siguientes tipos de consultas cuando se agota el tiempo de espera:

- Consultas de solo lectura, como las instrucciones SELECT, cuando tienen el estado de WLM `running`. Para encontrar el estado de WLM de una consulta, vea la columna STATE en la tabla de sistema STV_WLM_QUERY_STATE.
- Instrucciones CREATE TABLE AS (CTAS). El salto de colas de WLM es compatible con instrucciones CTAS definidas por el usuario y generadas por el sistema.
- Instrucciones SELECT INTO

Las consultas que no están sujetas al tiempo de espera de WLM siguen ejecutándose en la cola original hasta su finalización. Los siguientes tipos de consultas no están sujetos al tiempo de espera de WLM:

- Instrucciones COPY.
- Operaciones de mantenimiento, como ALTER, ANALYZE y VACUUM
- Consultas de solo lectura, como instrucciones SELECT, que han alcanzado el estado de WLM `returning`. Para encontrar el estado de WLM de una consulta, vea la columna STATE en la tabla de sistema STV_WLM_QUERY_STATE.

Las consultas que no pueden transferirse al agotar el tiempo de espera de WLM se cancelan. Los siguientes tipos de consultas no pueden transferirse al agotarse el tiempo de espera de WLM:

- Instrucciones INSERT, UPDATE y DELETE.
- Instrucciones UNLOAD.
- Funciones definidas por el usuario (UDF)

## Consultas reasignadas o reiniciadas por un tiempo de espera de WLM.

Cuando se transfiere una consulta y no se encuentra ninguna cola que coincida, la consulta se cancela.

Cuando se transfiere una consulta y se encuentra una cola coincidente, WLM intenta reasignar la consulta a esta nueva cola. Si la consulta no se puede reasignar, se reinicia en la nueva cola, tal y como se describe a continuación.

Una consulta solo puede reasignarse si se dan todas las condiciones siguientes:

- Se encuentra una cola coincidente.
- La nueva cola tiene suficientes slots libres para ejecutar la consulta. Una consulta puede necesitar varios slots si el parámetro wlm_query_slot_count se ha establecido en un valor mayor que 1.
- La nueva consulta tiene al menos tanta memoria disponible como la que utiliza actualmente.

Si la consulta se reasigna, sigue ejecutándose en la nueva cola. Los resultados intermedios se conservan, por lo que el efecto en el tiempo total de ejecución es mínimo.

Si la consulta no se puede reasignar, se cancela y se reinicia en la nueva cola. Los resultados intermedios se eliminan. La consulta espera en la cola y, cuando hay suficientes slots disponibles, comienza a ejecutarse.

## Acciones de transferencia de QMR

En la siguiente tabla, se resume el comportamiento de diferentes tipos de consulta con una acción de transferencia de QMR.

| Tipo de consulta | Acción |
| --- | --- |
| COPY | Siguen ejecutándose |
| Operaciones de mantenimiento | Siguen ejecutándose |
| Funciones definidas por el usuario (UDF) | Siguen ejecutándose |
| UNLOAD | Se reasignan o siguen ejecutándose |
| INSERT, UPDATE y DELETE | Se reasignan o siguen ejecutándose |
| Consultas de solo lectura con el estado `returning` | Se reasignan o siguen ejecutándose |
| Consultas de solo lectura con el estado `running` | Se reasignan o reinicia |
| CREATE TABLE AS (CTAS), SELECT INTO | Se reasignan o reinicia |

Para saber si una consulta transferida por QMR se ha reasignado, reiniciado o cancelado, consulte la tabla de registro del sistema STL_WLM_RULE_ACTION.

## Consultas reasignadas o reiniciadas por una acción de salto de QMR

Cuando se transfiere una consulta y no se encuentra ninguna cola que coincida, la consulta se cancela.

Cuando se transfiere una consulta y se encuentra una cola coincidente, WLM intenta reasignar la consulta a esta nueva cola. Si una consulta no se puede reasignar, se reinicia en la nueva cola o sigue ejecutándose en la cola original, tal y como se describe a continuación.

Una consulta solo puede reasignarse si se dan todas las condiciones siguientes:

- Se encuentra una cola coincidente.
- La nueva cola tiene suficientes slots libres para ejecutar la consulta. Una consulta puede necesitar varios slots si el parámetro wlm_query_slot_count se ha establecido en un valor mayor que 1.
- La nueva consulta tiene al menos tanta memoria disponible como la que utiliza actualmente.

Si la consulta se reasigna, sigue ejecutándose en la nueva cola. Los resultados intermedios se conservan, por lo que el efecto en el tiempo total de ejecución es mínimo.

Si una consulta no se puede reasignar, se reinicia o sigue ejecutándose en la cola original. Cuando una consulta se reinicia, se cancela y se reinicia en la nueva cola. Los resultados intermedios se eliminan. La consulta espera en la cola y, cuando hay suficientes slots disponibles, comienza a ejecutarse.

# Tutorial: Configuración de colas de Administración de carga de trabajo (WLM) manual

Con Amazon Redshift, puede configurar colas de administración manual de cargas de trabajo (WLM) para priorizar y asignar recursos a distintos tipos de consultas y usuarios. Las colas de WLM manuales le permiten controlar la configuración de memoria y simultaneidad de colas específicas, lo que garantiza que las cargas de trabajo críticas reciban los recursos necesarios y, al mismo tiempo, evita que las consultas de baja prioridad monopolicen el sistema. En las siguientes secciones, se le guiará a través del proceso de creación y configuración de colas de WLM manuales en Amazon Redshift para cumplir con los requisitos de administración de la carga de trabajo.

## Descripción general

Recomendamos configurar la administración de cargas de trabajo (WLM) automática en Amazon Redshift. Para obtener más información acerca de la WLM automática, consulte Administración de la carga de trabajo. No obstante, si necesita múltiples colas de WLM, en este tutorial, se detalla el proceso de configuración de la administración de cargas de trabajo (WLM) manual en Amazon Redshift. Al configurar WLM manual, puede mejorar el rendimiento de consultas y la asignación de recursos en el clúster.

Amazon Redshift dirige las consultas de usuarios a las colas para su procesamiento. WLM define el modo en que esas consultas se dirigen a las colas. De forma predeterminada, Amazon Redshift tiene dos colas disponibles para consultas: una para superusuarios y una para usuarios. La cola de

superusuario no se puede configurar y solo puede procesar una consulta a la vez. Debe reservar esta cola únicamente para solucionar problemas. La cola de usuarios puede procesar hasta cinco consultas a la vez, pero, si lo necesita, puede configurar esto modificando el nivel de simultaneidad de la cola.

Cuando tiene varios usuarios ejecutando consultas en la base de datos, tal vez otra configuración le resulte más eficaz. Por ejemplo, si algunos usuarios ejecutan operaciones que consumen muchos recursos, como VACUUM, esto puede tener un impacto negativo en las consultas que menos consumen, como informes. Podría plantearse agregar colas adicionales y configurarlas para cargas de trabajo diferentes.

Tiempo estimado: 75 minutos

Costo estimado: 50 céntimos

Requisitos previos

Necesita un clúster de Amazon Redshift, la base de datos de ejemplo TICKIT y la herramienta de cliente Amazon Redshift RSQL. Si todavía no tiene estos elementos configurados, vaya a Guía de introducción a Amazon Redshift y a Amazon Redshift RSQL.

Secciones

- Sección 1: Comportamiento del procesamiento de la cola predeterminada
- Sección 2: Modificación de la configuración de la cola de consultas de WLM
- Sección 3: Direccionamiento de consultas a las colas en función de los grupos de usuarios y de consultas
- Sección 4: Utilización de wlm_query_slot_count para anular de forma temporal el nivel de simultaneidad en una cola
- Sección 5: Limpieza de los recursos

## Sección 1: Comportamiento del procesamiento de la cola predeterminada

Antes de comenzar a configurar la WLM manual, es útil entender el comportamiento predeterminado del procesamiento de las colas en Amazon Redshift. En esta sección, crea dos vistas de bases de datos que devuelvan la información de varias tablas de sistema. Luego, ejecuta algunas consultas de prueba para ver cómo se dirigen las consultas de forma predeterminada. Para obtener más información acerca de las tablas de sistema, consulte Referencia de las tablas y vistas de sistema.

### Paso 1: Creación de la vista WLM_QUEUE_STATE_VW

En este paso, creará una vista denominada WLM_QUEUE_STATE_VW. Esta vista devuelve la información de las siguientes tablas de sistema.

- STV_WLM_CLASSIFICATION_CONFIG
- STV_WLM_SERVICE_CLASS_CONFIG
- STV_WLM_SERVICE_CLASS_STATE

Esta vista se utiliza en todo el tutorial para supervisar lo que les ocurre a las colas después de modificar la configuración de WLM. En la siguiente tabla se describen los datos que devuelve la vista WLM_QUEUE_STATE_VW.

| Columna | Descripción |
|---|---|
| cola | El número asociado a la fila que representa una cola. El número de cola determina el orden de las colas en la base de datos. |
| description | Valor que describe si una cola está disponible solo para algunos grupos de usuarios, algunos grupos de consultas o todos los tipos de consultas. |
| slots | El número de slots asignados a la cola. |
| mem | La cantidad de memoria, en MB por slot, asignada a la cola. |
| max_execution_time | La cantidad de tiempo que una consulta tiene permitido ejecutarse antes de terminarse. |
| user_* | Valor que indica si los caracteres comodines pueden coincidir con los grupos de usuarios en la configuración de WLM. |
| query_* | Valor que indica si los caracteres comodines pueden coincidir con los grupos de consultas en la configuración de WLM. |
| queued | El número de consultas que están en espera en la cola para ser procesadas. |
| executing | La cantidad de consultas que se encuentran en ejecución. |

| Columna | Descripción |
| --- | --- |
| executed | La cantidad de consultas que se han ejecutado. |

Creación de la vista WLM_QUEUE_STATE_VW

1. Abra Amazon Redshift RSQL y conéctese a la base de datos de ejemplo de TICKIT. Si no dispone de esta base de datos, consulte Requisitos previos.
2. Ejecute la siguiente consulta para crear la vista WLM_QUEUE_STATE_VW.

```
create view WLM_QUEUE_STATE_VW as
select (config.service_class-5) as queue
, trim (class.condition) as description
, config.num_query_tasks as slots
, config.query_working_mem as mem
, config.max_execution_time as max_time
, config.user_group_wild_card as "user_*"
, config.query_group_wild_card as "query_*"
, state.num_queued_queries queued
, state.num_executing_queries executing
, state.num_executed_queries executed
from
STV_WLM_CLASSIFICATION_CONFIG class,
STV_WLM_SERVICE_CLASS_CONFIG config,
STV_WLM_SERVICE_CLASS_STATE state
where
class.action_service_class = config.service_class
and class.action_service_class = state.service_class
and config.service_class > 4
order by config.service_class;
```

3. Ejecute la siguiente consulta para ver la información que contiene la vista.

```
select * from wlm_queue_state_vw;
```

A continuación se muestra un resultado de ejemplo.

```
query | description                             | slots | mem | max_time | user_* |
 query_* | queued | executing | executed
```

```
------+----------------------------------------------+-------+-----+----------+--------
+---------+--------+-----------+----------
    0 | (super user) and (query group: superuser) |     1 | 357 |        0 |  false |
 false   |      0 |         0 |        0
    1 | (querytype:any)                           |     5 | 836 |        0 |  false |
 false   |      0 |         1 |      160
```

Paso 2: Creación de la vista WLM_QUERY_STATE_VW

En este paso, crea una vista denominada WLM_QUERY_STATE_VW. Esta vista devuelve la información de la tabla de sistema STV_WLM_QUERY_STATE.

Esta vista se utiliza en todo el tutorial para supervisar las consultas que están en ejecución. En la siguiente tabla se describen los datos que devuelve la vista WLM_QUERY_STATE_VW.

| Columna | Descripción |
|---|---|
| consulta | El ID de la consulta. |
| cola | El número de cola. |
| slot_count | El número de slots asignados a la consulta. |
| start_time | La hora en que comenzó la consulta. |
| estado | El estado de la consulta, como executing (en ejecución). |
| queue_time | El número de microsegundos que la consulta estuvo en la cola. |
| exec_time | La cantidad de microsegundos que estuvo la consulta en la cola. |

Creación de la vista WLM_QUERY_STATE_VW

1. En RSQL, ejecute la siguiente consulta para crear la vista WLM_QUERY_STATE_VW.

```
create view WLM_QUERY_STATE_VW as
select query, (service_class-5) as queue, slot_count, trim(wlm_start_time) as
 start_time, trim(state) as state, trim(queue_time) as queue_time, trim(exec_time) as
 exec_time
from stv_wlm_query_state;
```

2. Ejecute la siguiente consulta para ver la información que contiene la vista.

```
select * from wlm_query_state_vw;
```

A continuación se muestra un resultado de ejemplo.

```
query | queue | slot_count | start_time          | state     | queue_time | exec_time
------+-------+------------+---------------------+-----------+------------
+-----------
 1249 |     1 |          1 | 2014-09-24 22:19:16 | Executing | 0          | 516
```

### Paso 3: Ejecución de consultas de prueba

En este paso, ejecuta consultas desde varias conexiones en RSQL y revisa las tablas de sistema para determinar cómo se dirigieron las consultas para su procesamiento.

Para este paso, necesita tener dos ventanas RSQL abiertas:

- En la ventana RSQL 1, se ejecutan consultas que monitoreen el estado de las colas y consultas que utilicen las vistas ya creadas en este tutorial.
- En la ventana RSQL 2, se pueden ejecutar consultas de ejecución prolongada para cambiar los resultados que se encuentren en la ventana 1 de RSQL.

#### Ejecución de las consultas de prueba

1. Abra las dos ventanas RSQL. Si ya tiene una ventana abierta, solo necesita abrir una segunda. Puede utilizar la misma cuenta de usuario para ambas conexiones.
2. En la ventana RSQL 1, ejecute la siguiente consulta.

```
select * from wlm_query_state_vw;
```

A continuación se muestra un resultado de ejemplo.

```
query | queue | slot_count | start_time          | state     | queue_time | exec_time
------+-------+------------+---------------------+-----------+------------
+-----------
 1258 |     1 |          1 | 2014-09-24 22:21:03 | Executing | 0          | 549
```

Esta consulta devuelve un resultado autorreferencial. La consulta que se encuentra en ejecución es la instrucción SELECT de esta vista. Una consulta en esta vista siempre devuelve como mínimo un resultado. Compara este resultado con el resultado que se produzca después de comenzar la consulta de ejecución prolongada, en el siguiente paso.

3. En la ventana RSQL 2, ejecute una consulta desde la base de datos de ejemplo TICKIT. Esta consulta se debe ejecutar por un minuto aproximadamente, de modo que tenga tiempo para explorar los resultados de las vistas WLM_QUEUE_STATE_VW y WLM_QUERY_STATE_VW que creó antes. En algunos casos, es posible que encuentre que la consulta no se ejecuta durante el tiempo necesario para consultar ambas vistas. En estos casos, puede aumentar el valor del filtro en `l.listid` para que se ejecute durante más tiempo.

   > **Note**
   >
   > Para reducir el tiempo de ejecución de las consultas y mejorar el rendimiento del sistema, Amazon Redshift almacena en caché los resultados de ciertos tipos de consultas en la memoria del nodo principal. Si se ha habilitado el almacenamiento en caché de los resultados, las consultas posteriores se ejecutarán mucho más rápido. Para evitar que la consulta se ejecute con demasiada rapidez, deshabilite el almacenamiento en caché de los resultados para la sesión actual.

   Para deshabilitar el almacenamiento en caché de los resultados para la sesión actual, establezca el parámetro enable_result_cache_for_session en `off`, tal y como se muestra a continuación.

   ```
   set enable_result_cache_for_session to off;
   ```

   En la ventana RSQL 2, ejecute la siguiente consulta.

   ```
   select avg(l.priceperticket*s.qtysold) from listing l, sales s where l.listid <
    100000;
   ```

4. En la ventana RSQL 1, consulte WLM_QUEUE_STATE_VW y WLM_QUERY_STATE_VW y compare los resultados con los resultados anteriores.

   ```
   select * from wlm_queue_state_vw;
   select * from wlm_query_state_vw;
   ```

A continuación se incluyen resultados de ejemplo.

```
query | description                                | slots | mem | max_time | user_* |
 query_* | queued | executing | executed
------+--------------------------------------------+-------+-----+----------+--------
+---------+--------+-----------+----------
    0 | (super user) and (query group: superuser) |     1 | 357 |        0 |  false |
 false   |      0 |         0 |        0
    1 | (querytype:any)                           |     5 | 836 |        0 |  false |
 false   |      0 |         2 |      163

query | queue | slot_count | start_time          | state     | queue_time | exec_time
------+-------+------------+---------------------+-----------+------------
+-----------
 1267 |     1 |          1 | 2014-09-24 22:22:30 | Executing | 0          | 684
 1265 |     1 |          1 | 2014-09-24 22:22:36 | Executing | 0          | 4080859
```

Observe las siguientes diferencias entre las consultas anteriores y los resultados de este paso:

- Ahora existen dos filas en WLM_QUERY_STATE_VW. Un resultado es la consulta autorreferencial por la ejecución de una operación SELECT en esta vista. El segundo resultado es la consulta de ejecución prolongada del paso anterior.
- La columna executing en WLM_QUEUE_STATE_VW aumentó de 1 a 2. Esta entrada de la columna significa que hay dos consultas en ejecución en la cola.
- La columna executed se incrementa cada vez que ejecute una consulta en la cola.

La vista WLM_QUEUE_STATE_VW es útil para obtener una vista general de las colas y saber cuántas consultas se están procesando en cada cola. La vista WLM_QUERY_STATE_VW es útil para obtener una vista más detallada de las consultas individuales que se están ejecutando.

## Sección 2: Modificación de la configuración de la cola de consultas de WLM

Ahora que ya conoce el funcionamiento predeterminado de las colas, puede aprender a configurar las colas de consultas con la WLM manual. En esta sección, crea y configura un nuevo grupo de parámetros para el clúster. Va a crear dos colas de usuario adicionales y las va a configurar para que acepten consultas en función de las etiquetas del grupo de consultas o del grupo de usuarios de las consultas. Cualquier consulta que no se dirija a una de estas dos colas se dirigirá a la cola predeterminada en tiempo de ejecución.

Para crear una configuración manual de WLM en un grupo de parámetros

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. En el menú de navegación, elija Configurations (Configuraciones) y, a continuación, elija Workload management (Administración de la carga de trabajo) para mostrar la página de Workload management (Administración de la carga de trabajo).
3. Seleccione Create (Crear) para mostrar la ventana de Create parameter group (Crear grupo de parámetros).
4. Ingrese **`WLMTutorial`** en Parameter group name (Nombre del grupo de parámetros) y Description (Descripción), y luego elija Create (Crear) para generar el grupo de parámetros.

   > **Note**
   >
   > El nombre del grupo de parámetros se convierte a un formato de todo en minúsculas cuando se crea.

5. En la página de Workload management (Administración de la carga de trabajo), seleccione el grupo de parámetros **`wlmtutorial`** para mostrar la página de detalles con pestañas para Parameters (Parámetros) y Workload management (Administración de la carga de trabajo).
6. Asegúrese de que está en la pestaña Workload management (Administración de la carga de trabajo) y luego seleccione Switch WLM mode (Cambiar modo de WLM) para mostrar la ventana de Concurrency settings (Configuración de concurrencia).
7. Seleccione WLM manual y luego Save (Guardar) para cambiar al WLM manual.
8. Seleccione Edit workload queues (Editar colas de carga de trabajo).
9. Seleccione dos veces Add queue (Agregar cola) para agregar dos colas. Ahora hay tres colas: Cola 1, Cola 2 y Cola predeterminada.
10. Introduzca información para cada cola como se indica a continuación:

    - En la Cola 1, introduzca **`30`** para Memory (Memoria) (%), **`2`** para Concurrency on main (Concurrencia en principal) y **`test`** para Query groups (Grupos de consulta). No cambie los demás valores predeterminados.
    - En la Cola 2, introduzca **`40`** para Memory (Memoria) (%), **`3`** para Concurrency on main (Concurrencia en principal) y **`admin`** para User groups (Grupos de usuario). No cambie los demás valores predeterminados.

- Establezca el valor de Simultaneidad en el principal de la cola predeterminada en un valor mayor o igual a 1. No realice ningún cambio en Cola predeterminada. WLM asigna la memoria sin asignar a la cola predeterminada.

11. Seleccione Save (Guardar) para guardar las configuraciones.

A continuación, asocie el grupo de parámetros que tiene la configuración manual de WLM con un clúster.

Para asociar un grupo de parámetros que tenga una configuración manual de WLM con un clúster

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en https://console.aws.amazon.com/redshiftv2/.
2. En el menú de navegación, elija Clusters (Clústeres) y, a continuación, elija Clusters (Clústeres) para mostrar una lista de los clústeres.
3. Elija el clúster, como `examplecluster`, para mostrar los detalles del clúster. A continuación, elija la pestaña Properties (Propiedades) para mostrar las propiedades de ese clúster.
4. En la sección Database configurations (Configuración de las bases de datos), elija Edit (Editar) y Edit parameter group (Editar grupo de parámetros) para mostrar la ventana de grupos de parámetros.
5. En Parameter groups (Grupos de parámetros), elija el grupo de parámetros **`wlmtutorial`** creado anteriormente.
6. Elija Save changes (Guardar cambios) para asociar el grupo de parámetros.

   El clúster se modifica con el grupo de parámetros cambiado. No obstante, tendrá que reiniciar el cluster para que se apliquen los cambios en la base de datos.
7. Elija su clúster y luego elija Reboot (Reiniciar) en Actions (Acciones).

Una vez que se reinicie el clúster, volverá al estado Available (Disponible).

## Sección 3: Direccionamiento de consultas a las colas en función de los grupos de usuarios y de consultas

Ahora tiene el clúster asociado un grupo de parámetros nuevo y ha configurado WLM. A continuación, ejecute algunas consultas para ver cómo Amazon Redshift dirige las consultas a las colas para su procesamiento.

### Paso 1: Vista de la configuración de la cola de consultas en la base de datos

Primero, verifique que la base de datos tenga la configuración WLM que espera.

Para ver la configuración de la cola de consultas

1. Abra RSQL y ejecute la siguiente consulta. La consulta utiliza la vista WLM_QUEUE_STATE_VW que el usuario creó en Paso 1: Creación de la vista WLM_QUEUE_STATE_VW. Si ya tenía una sesión conectada a la base de datos antes de reiniciar el clúster, deberá volver a conectarla.

   ```
   select * from wlm_queue_state_vw;
   ```

   A continuación se muestra un resultado de ejemplo.

   ```
   query | description                                | slots | mem | max_time | user_* |
    query_* | queued | executing | executed
   ------+--------------------------------------------+-------+-----+----------+--------
   +---------+--------+-----------+----------
       0 | (super user) and (query group: superuser) |     1 | 357 |        0 |  false |
    false   |      0 |         0 |        0
       1 | (query group: test)                       |     2 | 627 |        0 |  false |
    false   |      0 |         0 |        0
       2 | (suser group: admin)                      |     3 | 557 |        0 |  false |
    false   |      0 |         0 |        0
       3 | (querytype:any)                           |     5 | 250 |        0 |  false |
    false   |      0 |         1 |        0
   ```

   Compare estos resultados con los resultados recibidos en Paso 1: Creación de la vista WLM_QUEUE_STATE_VW. Observe que ahora hay dos colas adicionales. La cola 1 ahora es la cola para el grupo de consultas de prueba y la cola 2 es la cola para el grupo del usuario administrador.

   La cola 3 ahora es la cola predeterminada. La última cola de la lista siempre es la cola predeterminada. Esa es la cola a la que se dirigen las consultas de forma predeterminada si no se especifica ningún grupo de usuarios o de consultas en una consulta.
2. Ejecute la siguiente consulta para confirmar que la consulta ahora se ejecuta en la cola 3.

   ```
   select * from wlm_query_state_vw;
   ```

   A continuación se muestra un resultado de ejemplo.

```
query | queue | slot_count | start_time          | state     | queue_time | exec_time
------+-------+------------+---------------------+-----------+------------
+-----------
 2144 |     3 |          1 | 2014-09-24 23:49:59 | Executing | 0          | 550430
```

Paso 2: Ejecución de una consulta mediante la cola de grupo de consultas

Para ejecutar una consulta mediante la cola de grupo de consultas

1. Ejecute la siguiente consulta para dirigirla al grupo de consultas `test`.

```
set query_group to test;
select avg(l.priceperticket*s.qtysold) from listing l, sales s where l.listid <40000;
```

2. Desde la otra ventana RSQL, ejecute la siguiente consulta.

```
select * from wlm_query_state_vw;
```

   A continuación se muestra un resultado de ejemplo.

```
query | queue | slot_count | start_time          | state     | queue_time | exec_time
------+-------+------------+---------------------+-----------+------------
+-----------
 2168 |     1 |          1 | 2014-09-24 23:54:18 | Executing | 0          | 6343309
 2170 |     3 |          1 | 2014-09-24 23:54:24 | Executing | 0          | 847
```

   La consulta se dirigió al grupo de consultas de prueba, que ahora es la cola 1.

3. Seleccione todos los elementos de la vista de estado de la cola.

```
select * from wlm_queue_state_vw;
```

   Se ve un resultado similar al siguiente.

```
query | description                                | slots | mem | max_time | user_* |
 query_* | queued | executing | executed
------+--------------------------------------------+-------+-----+----------+--------
+---------+--------+-----------+----------
```

```
    0 | (super user) and (query group: superuser) |     1 | 357 |        0 |  false |
 false   |      0 |         0 |        0
    1 | (query group: test)                       |     2 | 627 |        0 |  false |
 false   |      0 |         1 |        0
    2 | (suser group: admin)                      |     3 | 557 |        0 |  false |
 false   |      0 |         0 |        0
    3 | (querytype:any)                           |     5 | 250 |        0 |  false |
 false   |      0 |         1 |        0
```

4. Ahora, restablezca el grupo de consultas y vuelva a ejecutar la consulta larga:

```
reset query_group;
select avg(l.priceperticket*s.qtysold) from listing l, sales s where l.listid <40000;
```

5. Ejecute las consultas en las vistas para ver los resultados.

```
select * from wlm_queue_state_vw;
select * from wlm_query_state_vw;
```

   A continuación se incluyen resultados de ejemplo.

```
query | description                               | slots | mem | max_time | user_* |
 query_* | queued | executing | executed
------+-------------------------------------------+-------+-----+----------+--------
+---------+--------+-----------+----------
    0 | (super user) and (query group: superuser) |     1 | 357 |        0 |  false |
 false   |      0 |         0 |        0
    1 | (query group: test)                       |     2 | 627 |        0 |  false |
 false   |      0 |         0 |        1
    2 | (suser group: admin)                      |     3 | 557 |        0 |  false |
 false   |      0 |         0 |        0
    3 | (querytype:any)                           |     5 | 250 |        0 |  false |
 false   |      0 |         2 |        5

query | queue | slot_count | start_time          | state     | queue_time | exec_time
------+-------+------------+---------------------+-----------+------------
+-----------
 2186 |     3 |          1 | 2014-09-24 23:57:52 | Executing | 0          | 649
 2184 |     3 |          1 | 2014-09-24 23:57:48 | Executing | 0          | 4137349
```

   El resultado debe ser que la consulta ahora se ejecute en la cola 3 nuevamente.

### Paso 3: Creación de un usuario y un grupo de base de datos

Antes de poder ejecutar cualquier consulta en esta cola, debe crear el grupo de usuarios en la base de datos y agregar un usuario al grupo. A continuación, inicie sesión con RSQL con las nuevas credenciales de usuario y ejecute las consultas. Debe ejecutar las consultas como superusuario, como por ejemplo, usuario administrador, para crear usuarios de bases de datos.

Para crear un usuario y un grupo de usuarios de base de datos nuevos

1. En la base de datos, cree un usuario de base de datos nuevo denominado `adminwlm` mediante la ejecución del siguiente comando en una ventana RSQL.

```
create user adminwlm createuser password '123Admin';
```

2. Luego, ejecute los siguientes comandos para crear un nuevo grupo de usuarios y agregue el nuevo usuario `adminwlm` a él.

```
create group admin;
alter group admin add user adminwlm;
```

### Paso 4: Ejecución de una consulta mediante la cola de grupo de usuarios

A continuación, ejecuta una consulta y la dirige a la cola de grupo de usuarios. Esto se hace cuando desea dirigir la consulta a una cola que esté configurada para administrar el tipo de consulta que desea ejecutar.

Para ejecutar una consulta mediante la cola de grupo de usuarios

1. En la ventana RSQL 2, ejecute las siguientes consultas para cambiar a la cuenta `adminwlm` y ejecute una consulta como ese usuario.

```
set session authorization 'adminwlm';
select avg(l.priceperticket*s.qtysold) from listing l, sales s where l.listid <40000;
```

2. En la ventana RSQL 1, ejecute la siguiente consulta para ver la cola de consultas a la que se dirigen las consultas.

```
select * from wlm_query_state_vw;
select * from wlm_queue_state_vw;
```

A continuación se incluyen resultados de ejemplo.

```
query | description                                | slots | mem | max_time | user_* |
 query_* | queued | executing | executed
------+--------------------------------------------+-------+-----+----------+--------
+---------+--------+-----------+----------
    0 | (super user) and (query group: superuser) |     1 | 357 |        0 |  false |
 false   |      0 |         0 |        0
    1 | (query group: test)                       |     2 | 627 |        0 |  false |
 false   |      0 |         0 |        1
    2 | (suser group: admin)                      |     3 | 557 |        0 |  false |
 false   |      0 |         1 |        0
    3 | (querytype:any)                           |     5 | 250 |        0 |  false |
 false   |      0 |         1 |        8

query | queue | slot_count | start_time          | state     | queue_time | exec_time
------+-------+------------+---------------------+-----------+------------
+-----------
 2202 |     2 |          1 | 2014-09-25 00:01:38 | Executing | 0          | 4885796
 2204 |     3 |          1 | 2014-09-25 00:01:43 | Executing | 0          | 650
```

La cola en la que se ejecuta esta consulta es la cola 2, la cola de usuario `admin`. Cada vez que inicie sesión como este usuario y ejecute consultas, estas se ejecutan en la cola 2 a menos que especifique otro grupo de consultas para utilizar. La cola elegida depende de las reglas de asignación de colas. Para obtener más información, consulte Reglas de asignación de colas de WLM.

3. Ahora ejecute la siguiente consulta desde la ventana RSQL 2.

```
set query_group to test;
select avg(l.priceperticket*s.qtysold) from listing l, sales s where l.listid <40000;
```

4. En la ventana RSQL 1, ejecute la siguiente consulta para ver la cola de consultas a la que se dirigen las consultas.

```
select * from wlm_queue_state_vw;
select * from wlm_query_state_vw;
```

A continuación se incluyen resultados de ejemplo.

```
query | description                                | slots | mem | max_time | user_* |
 query_* | queued | executing | executed
------+--------------------------------------------+-------+-----+----------+--------
+---------+--------+-----------+----------
    0 | (super user) and (query group: superuser) |     1 | 357 |        0 |  false |
 false   |      0 |         0 |        0
    1 | (query group: test)                       |     2 | 627 |        0 |  false |
 false   |      0 |         1 |        1
    2 | (suser group: admin)                      |     3 | 557 |        0 |  false |
 false   |      0 |         0 |        1
    3 | (querytype:any)                           |     5 | 250 |        0 |  false |
 false   |      0 |         1 |       10

query | queue | slot_count | start_time          | state     | queue_time | exec_time
------+-------+------------+---------------------+-----------+------------
+-----------
 2218 |     1 |          1 | 2014-09-25 00:04:30 | Executing | 0          | 4819666
 2220 |     3 |          1 | 2014-09-25 00:04:35 | Executing | 0          | 685
```

5. Cuando haya terminado, restablezca el grupo de consultas.

```
reset query_group;
```

## Sección 4: Utilización de wlm_query_slot_count para anular de forma temporal el nivel de simultaneidad en una cola

En ocasiones, los usuarios pueden necesitar de forma temporal más recursos para una consulta en particular. De ser así, pueden utilizar la opción de configuración wlm_query_slot_count para anular temporalmente la forma en la que se asignan los slots en una cola de consultas. Los slots son unidades de memoria y CPU que se utilizan para procesar consultas. Puede anular el número de slots cuando tiene consultas ocasionales que consumen una gran cantidad de recursos del clúster, como cuando realiza una operación VACUUM en la base de datos.

Es posible que se encuentre con que los usuarios necesitan con frecuencia establecer para determinados tipos de consulta. Si es así, plantéese ajustar la configuración de WLM y proporcionar a los usuarios una cola que se adapte mejor a las necesidades de sus consultas. Para obtener más información acerca de la anulación temporal del nivel de simultaneidad utilizando el número de slots, consulte wlm_query_slot_count.

### Paso 1: Anulación del nivel de simultaneidad mediante wlm_query_slot_count

Para los fines de este tutorial, ejecutamos la misma consulta de ejecución prolongada SELECT. La ejecutamos como el usuario `adminwlm` con wlm_query_slot_count para aumentar el número de slots disponibles para la consulta.

Para anular el nivel de simultaneidad mediante wlm_query_slot_count

1. Aumente el límite en la consulta para asegurarse de que dispone del tiempo suficiente para consultar la vista WLM_QUERY_STATE_VW y ver un resultado.

   ```
   set wlm_query_slot_count to 3;
   select avg(l.priceperticket*s.qtysold) from listing l, sales s where l.listid <40000;
   ```

2. Ahora, la consulta WLM_QUERY_STATE_VW utiliza el usuario administrador para ver cómo se ejecuta la consulta.

   ```
   select * from wlm_query_state_vw;
   ```

   A continuación se muestra un resultado de ejemplo.

   ```
   query | queue | slot_count | start_time          | state     | queue_time | exec_time
   ------+-------+------------+---------------------+-----------+------------
   +-----------
    2240 |     2 |          1 | 2014-09-25 00:08:45 | Executing | 0          | 3731414
    2242 |     3 |          1 | 2014-09-25 00:08:49 | Executing | 0          | 596
   ```

   Observe que el número de slots para la consulta es 3. Este número significa que la consulta está utilizando los tres slots para procesarse, lo que lleva a la asignación de todos los recursos de la cola a esa consulta.

3. Ahora, ejecute la siguiente consulta.

   ```
   select * from WLM_QUEUE_STATE_VW;
   ```

   A continuación se muestra un resultado de ejemplo.

   ```
   query | description                                | slots | mem | max_time | user_* |
    query_* | queued | executing | executed
   ------+--------------------------------------------+-------+-----+----------+--------
   +---------+--------+-----------+----------
   ```

```
    0 | (super user) and (query group: superuser) |     1 | 357 |        0 |  false |
 false   |      0 |         0 |        0
    1 | (query group: test)                       |     2 | 627 |        0 |  false |
 false   |      0 |         0 |        4
    2 | (suser group: admin)                      |     3 | 557 |        0 |  false |
 false   |      0 |         1 |        3
    3 | (querytype:any)                           |     5 | 250 |        0 |  false |
 false   |      0 |         1 |       25
```

La opción de configuración wlm_query_slot_count es válida solo para la sesión actual. Si esa sesión expira u otro usuario ejecuta una consulta, se utiliza la configuración de WLM.

4. Restablezca el número de slots y vuelva a ejecutar la prueba.

```
reset wlm_query_slot_count;
select avg(l.priceperticket*s.qtysold) from listing l, sales s where l.listid <40000;
```

A continuación se incluyen resultados de ejemplo.

```
query | description                               | slots | mem | max_time | user_* |
 query_* | queued | executing | executed
------+-------------------------------------------+-------+-----+----------+--------
+---------+--------+-----------+----------
    0 | (super user) and (query group: superuser) |     1 | 357 |        0 |  false |
 false   |      0 |         0 |        0
    1 | (query group: test)                       |     2 | 627 |        0 |  false |
 false   |      0 |         0 |        2
    2 | (suser group: admin)                      |     3 | 557 |        0 |  false |
 false   |      0 |         1 |        2
    3 | (querytype:any)                           |     5 | 250 |        0 |  false |
 false   |      0 |         1 |       14

query | queue | slot_count | start_time          | state     | queue_time | exec_time
------+-------+------------+---------------------+-----------+------------
+-----------
 2260 |     2 |          1 | 2014-09-25 00:12:11 | Executing | 0          | 4042618
 2262 |     3 |          1 | 2014-09-25 00:12:15 | Executing | 0          | 680
```

## Paso 2: Ejecución de consultas de sesiones diferentes

A continuación, ejecute consultas de sesiones diferentes.

Para ejecutar consultas de sesiones diferentes

1. En las ventanas RSQL 1 y 2, ejecute lo siguiente para utilizar el grupo de consultas de prueba.

```
set query_group to test;
```

2. En la ventana RSQL 1, ejecute la siguiente consulta de ejecución prolongada.

```
select avg(l.priceperticket*s.qtysold) from listing l, sales s where l.listid <40000;
```

3. Mientras se ejecuta la consulta de ejecución prolongada en la ventana RSQL 1, ejecute lo siguiente. Estos comandos aumentan el recuento de slots para usar todos los slots para la cola y empezar a ejecutar la consulta de ejecución prolongada a continuación.

```
set wlm_query_slot_count to 2;
select avg(l.priceperticket*s.qtysold) from listing l, sales s where l.listid <40000;
```

4. Abra una tercera ventana RSQL y consulte las vistas para ver los resultados.

```
select * from wlm_queue_state_vw;
select * from wlm_query_state_vw;
```

   A continuación se incluyen resultados de ejemplo.

```
query | description                                | slots | mem | max_time | user_* |
 query_* | queued | executing | executed
------+--------------------------------------------+-------+-----+----------+--------
+---------+--------+-----------+----------
    0 | (super user) and (query group: superuser) |     1 | 357 |        0 |  false |
 false   |      0 |         0 |        0
    1 | (query group: test)                       |     2 | 627 |        0 |  false |
 false   |      1 |         1 |        2
    2 | (suser group: admin)                      |     3 | 557 |        0 |  false |
 false   |      0 |         0 |        3
    3 | (querytype:any)                           |     5 | 250 |        0 |  false |
 false   |      0 |         1 |       18

query | queue | slot_count | start_time          | state         | queue_time |
 exec_time
------+-------+------------+---------------------+---------------+------------
+-----------
 2286 |     1 |          2 | 2014-09-25 00:16:48 | QueuedWaiting | 3758950    | 0
```

```
2282 |     1 |          1 | 2014-09-25 00:16:33 | Executing     | 0          |
19335850
2288 |     3 |          1 | 2014-09-25 00:16:52 | Executing     | 0          | 666
```

Observe que la primera consulta utiliza uno de los slots asignados a la cola 1 para ejecutar la consulta. Observe, además, que hay un consulta esperando en la cola (donde `queued` es `1` y `state` es `QueuedWaiting`). Una vez que se completa la primera consulta, la segunda comienza a ejecutarse. Esta ejecución ocurre porque ambas consultas se dirigen al grupo de consultas `test` y la segunda consulta debe esperar por la cantidad de slots suficientes para comenzar a procesarse.

### Sección 5: Limpieza de los recursos

El clúster seguirá acumulando cargos mientras esté en ejecución. Una vez que haya completado este tutorial, restablezca el entorno al estado anterior y, para ello, siga los pasos de Búsqueda de recursos adicionales y restablecimiento del entorno en la Guía de introducción a Amazon Redshift.

Para obtener más información acerca de WLM, consulte Administración de la carga de trabajo.

## Escalado de simultaneidad

Con la característica de escalado de simultaneidad, puede admitir miles de consultas y usuarios simultáneos con un rendimiento rápido de las consultas en todo momento. Cuando activa el escalado de simultaneidad, Amazon Redshift agrega capacidad al clúster de manera automática para procesar un aumento de las consultas de lectura y escritura. Los usuarios ven siempre los datos más actualizados, tanto si las consultas se ejecutan en el clúster principal como si se ejecutan en un clúster de escalado de simultaneidad.

Puede administrar qué consultas se envían al clúster de escalado de simultaneidad a través de la configuración de colas de WLM. Cuando activa el escalado de simultaneidad, las consultas que cumplen los requisitos necesarios se envían al clúster de escalado de simultaneidad en lugar de esperar en una cola.

Solo se le cobrará el tiempo durante el que los clústeres de escalado de simultaneidad estén ejecutando consultas. Para obtener más información sobre los precios, incluido el modo en que se reparten los cargos y los cargos mínimos, consulte Precios de escalado simultáneo.

Temas

- Capacidades de escalado de simultaneidad
- Limitaciones para el escalado de simultaneidad
- Regiones de AWS para el escalado de simultaneidad
- Candidatos al escalado de simultaneidad
- Configuración de colas de escalado de simultaneidad
- Monitoreo del escalado de simultaneidad
- Vistas del sistema de escalado de simultaneidad


## Capacidades de escalado de simultaneidad

Cuando activa el escalado de simultaneidad para una cola de WLM, funciona para operaciones de lectura, como consultas de paneles. También funciona para operaciones de escritura de uso común, como instrucciones para la ingesta y el procesamiento de datos.

### Capacidades de escalado de simultaneidad para las operaciones de escritura

El escalado de simultaneidad admite operaciones de escritura frecuentes, como instrucciones de extracción, transformación y carga (ETL). El escalado de simultaneidad para operaciones de escritura es especialmente útil si desea mantener tiempos de respuesta constantes cuando el clúster reciba un gran número de solicitudes. Mejora el rendimiento de las operaciones de escritura que compiten por los recursos en el clúster principal.

El escalado de simultaneidad admite las instrucciones COPY, INSERT, DELETE, UPDATE y CREATE TABLE AS (CTAS). Además, el escalado de simultaneidad admite la actualización manual de vistas materializadas (MV). No se admiten otras instrucciones de lenguaje de manipulación de datos (DML) ni de lenguaje de definición de datos (DDL). Si las instrucciones de escritura no compatibles, como CREATE sin TABLE AS, se incluyen en una transacción explícita antes de las instrucciones de escritura admitidas, ninguna de las instrucciones de escritura se ejecutará en clústeres escalados simultáneamente.

Cuando acumula crédito para el escalado de concurrencia, esta acumulación se aplica a las operaciones de lectura y escritura

## Limitaciones para el escalado de simultaneidad

A continuación, se detallan las limitaciones para utilizar el escalado de simultaneidad de Amazon Redshift:

- No admite consultas en tablas con claves de ordenación intercalada.
- No admite consultas en tablas temporales.
- No admite consultas que acceden a recursos externos protegidos por configuraciones de redes restrictivas o nubes virtuales privadas (VPC).
- No admite consultas que contengan funciones definidas por el usuario (UDF) de Python ni UDF de Lambda.
- No admite consultas que acceden a tablas de sistema, tablas de catálogo de PostgreSQL o tablas sin copias de seguridad.
- No admite consultas COPY o UNLOAD con acceso a un recurso externo cuando existen permisos de políticas de IAM restrictivas. Esto incluye permisos aplicados al recurso, como un bucket de Amazon S3 o tabla de DynamoDB, o al origen. Los orígenes de IAM pueden incluir los siguientes:
  - `aws:sourceVpc`: una VPC de origen.
  - `aws:sourceVpce`: un punto de conexión de VPC de origen.
  - `aws:sourceIp`: una dirección IP de origen.

  En algunos casos, es posible que necesite eliminar permisos que restringen el recurso o el origen, por lo que las consultas COPY y UNLOAD con acceso al recurso se envían al clúster de escalado de simultaneidad.

  Para obtener más información sobre las políticas de recursos, consulte Tipos de políticas en la guía del usuario de AWS Identity and Access Management y Cómo controlar el acceso desde los puntos de conexión de VPC con políticas de bucket.
- En las operaciones de DDL, como CREATE TABLE o ALTER TABLE, no se admite el escalado de simultaneidad de Amazon Redshift para operaciones de escritura.
- No admite ANALYZE para el comando COPY.
- No admite operaciones de escritura en una tabla de destino donde DISTSTYLE se configura como ALL.
- No admite COPY de los siguientes formatos de archivo:
  - Parquet
  - ORC
- No admite operaciones de escritura en tablas con columnas de identidad.
- Amazon Redshift admite el escalado de simultaneidad para operaciones de escritura solo en nodos RA3 de Amazon Redshift. El escalado de simultaneidad para operaciones de escritura no es compatible con otros tipos de nodos.

# Regiones de AWS para el escalado de simultaneidad

Con Amazon Redshift, puede utilizar el escalado simultáneo para administrar las demandas de carga de trabajo simultáneas en todos los clústeres de Redshift. En este tema se detalla en qué regiones puede utilizar el escalado simultáneo con Amazon Redshift.

El escalado de simultaneidad está disponible en estas regiones de AWS:

- Región Este de EE. UU. (Norte de Virginia) (us-east-1)
- Región EE. UU. Este (Ohio) (us-east-2)
- Región EE. UU. Oeste (Norte de California) (us-west-1)
- Región EE. UU. Oeste (Oregón) (us-west-2)
- Región Asia-Pacífico (Mumbai) (ap-south-1)
- Región Asia-Pacífico (Seúl) (ap-northeast-2)
- Región Asia-Pacífico (Singapur) (ap-southeast-1)
- Región Asia-Pacífico (Sídney) (ap-southeast-2)
- Región Asia-Pacífico (Malasia) (ap-southeast-5)
- Región Asia-Pacífico (Tokio) (ap-northeast-1)
- Región Canadá (Central) (ca-central-1)
- Región Oeste de Canadá (Calgary) (ca-west-1)
- Región China (Pekín) (cn-north-1)
- Región China (Ningxia) (cn-northwest-1)
- Región Europa (Fráncfort) (eu-central-1)
- Región Europa (Irlanda) (eu-west-1)
- Región Europa (Londres) (eu-west-2)
- Región Europa (París) (eu-west-3)
- Región Europa (Estocolmo) (eu-north-1)
- Región Europa (Zúrich) (eu-central-2)
- Región Europa (España) (eu-south-2)
- Región de Israel (Tel Aviv) (il-central-1)
- Región América del Sur (São Paulo) (sa-east-1)
- AWS GovCloud (Este de EE. UU.)

## Candidatos al escalado de simultaneidad

Con Amazon Redshift, puede escalar horizontalmente el procesamiento de consultas para acelerar la ejecución de consultas simultáneas. En el siguiente tema se describen los criterios que Amazon Redshift utiliza para determinar qué consultas se deben dirigir al escalado simultáneo.

Las consultas se envían al clúster de escalado de simultaneidad solo cuando el clúster principal cumple los siguientes requisitos:

- Plataforma EC2-VPC
- El nodo debe ser de tipo dc2.8xlarge, dc2.large, ra3.large, ra3.xlplus, ra3.4xlarge o ra3.16xlarge. El escalado de simultaneidad para operaciones de escritura solo se admite en nodos RA3 de Amazon Redshift.
- Máximo de 32 nodos de computación para clústeres con tipos de nodo ra3.xlplus, ra3.4xlarge o ra3.16xlarge. Además, el número de nodos del clúster principal no puede ser mayor de 32 nodos cuando se creó originalmente el clúster. Por ejemplo, si un clúster tiene actualmente 20 nodos pero se creó originalmente con 40, no cumple los requisitos de escalado de simultaneidad. Por el contrario, si un clúster de DC2 tiene actualmente 40 nodos pero se creó originalmente con 20, cumple los requisitos de escalado de simultaneidad.
- No es un clúster de un solo nodo.

## Configuración de colas de escalado de simultaneidad

Con Amazon Redshift, puede administrar la simultaneidad y los recursos del sistema mediante la configuración del escalado de simultaneidad. Las colas de escalado de simultaneidad le permiten establecer límites en el número de consultas o sesiones de usuario que se pueden ejecutar simultáneamente. En la siguiente sección se proporcionan instrucciones sobre cómo habilitar el escalado de simultaneidad de colas en Amazon Redshift, lo que le permite gestionar consultas y sesiones de usuario simultáneas de forma eficaz.

Las consultas se envían a los clústeres de escalado de simultaneidad al habilitar dicho escalado en una cola del administrador de cargas de trabajo (WLM). Para activar el escalado de simultaneidad en una cola, establezca el valor correspondiente a Concurrency Scaling mode (Modo de escalado de simultaneidad) en auto (automático).

Cuando el número de consultas enrutadas a una cola con el escalado de simultaneidad habilitado supera la capacidad de simultaneidad de la cola, tanto si la capacidad se configura de forma manual como si se determina de manera automática, las consultas aptas se envían al clúster de escalado de

simultaneidad. Cuando las ranuras de cola están disponibles en el clúster principal, las consultas se enrutan y ejecutan en el clúster principal. Como con cualquier cola WLM, las consultas se enrutan a una cola de escalado de simultaneidad basada en grupos de usuarios, etiquetando las consultas con etiquetas de grupos de consulta o según las condiciones de coincidencia definidas en Asignación de consultas a colas. También puede enviar las consultas definiendo Reglas de monitoreo de consultas de WLM. Por ejemplo, podría enviar todas las consultas que tarden más de cinco segundos a una cola de escalado de simultaneidad. Tenga en cuenta que el comportamiento de las colas puede variar, en función de si utiliza WLM automática o manual. Para obtener más información, consulte Implementación de WLM automática o Implementación de WLM manual.

El número predeterminado de clústeres de escalado de simultaneidad es uno. max_concurrency_scaling_clusters controla el número de clústeres de escalado de simultaneidad que se pueden utilizar.

## Monitoreo del escalado de simultaneidad

Con Amazon Redshift, puede supervisar y administrar el escalado de simultaneidad para optimizar el rendimiento y la rentabilidad de las cargas de trabajo de almacenamiento de datos. El escalado de simultaneidad permite a Amazon Redshift agregar automáticamente capacidad de clúster adicional cuando la demanda de carga de trabajo aumenta y eliminar esa capacidad cuando la demanda disminuye. En la sección siguiente, se proporciona orientación sobre cómo supervisar el escalado de simultaneidad de los clústeres de Amazon Redshift.

Para ver si se está ejecutando una consulta en el clúster principal o en un clúster de escalado de simultaneidad en la consola de Amazon Redshift, navegue hasta Clúster y elija un clúster. A continuación, seleccione la pestaña Monitoreo de consultas y Simultaneidad de cargas de trabajo para ver información sobre las consultas en ejecución y las consultas en cola.

Para buscar las horas de ejecución, consulte la tabla STL_QUERY y filtre por la columna `concurrency_scaling_status`. La siguiente consulta compara el tiempo de cola y el tiempo de ejecución de las consultas del clúster de escalado de simultaneidad y de las consultas que se ejecutan en el clúster principal.

```
SELECT w.service_class AS queue
, CASE WHEN q.concurrency_scaling_status = 1 THEN 'concurrency scaling cluster' ELSE
 'main cluster' END as concurrency_scaling_status
, COUNT( * ) AS queries
, SUM( q.aborted ) AS aborted
, SUM( ROUND( total_queue_time::NUMERIC / 1000000,2) ) AS queue_secs
, SUM( ROUND( total_exec_time::NUMERIC / 1000000,2) ) AS exec_secs
```

```
FROM stl_query q
JOIN stl_wlm_query w
USING (userid,query)
WHERE q.userid > 1
AND q.starttime > '2019-01-04 16:38:00'
AND q.endtime < '2019-01-04 17:40:00'
GROUP BY 1,2
ORDER BY 1,2;
```

Ajuste los valores `starttime` y `endtime` según los requisitos.

## Vistas del sistema de escalado de simultaneidad

Con Amazon Redshift, puede utilizar las vistas del sistema de escalado de simultaneidad para supervisar y administrar la actividad de escalado de simultaneidad en el clúster. En la siguiente sección, se describe la consulta de estas vistas del sistema y la interpretación de los resultados para aprovechar de forma eficaz el escalado simultáneo en el entorno de Amazon Redshift.

Un conjunto de vistas del sistema con el prefijo SVCS proporciona información de las tablas de registro del sistema sobre las consultas de los clústeres principal y de escalado de simultaneidad.

Las siguientes vistas contienen información similar a la correspondiente a las vistas STL o a las vistas SVL:

- SVCS_ALERT_EVENT_LOG
- SVCS_COMPILE
- SVCS_EXPLAIN
- SVCS_PLAN_INFO
- SVCS_QUERY_SUMMARY
- SVCS_STREAM_SEGS

Las siguientes vistas son específicas del escalado de simultaneidad.

- SVCS_CONCURRENCY_SCALING_USAGE

Para obtener más información acerca del escalado de simultaneidad, consulte los siguientes temas en la Guía de administración de Amazon Redshift.

- Consulta de datos de escalado de simultaneidad

- Visualización de rendimiento del clúster durante la ejecución de consultas
- Visualización de detalles de consultas

# Aceleración de consultas cortas

La aceleración de consultas cortas (SQA) da prioridad a una serie de consultas seleccionadas que se ejecutan rápidamente frente a consultas que tardan más en ejecutarse. SQA ejecuta las consultas cortas en un espacio dedicado, de forma que estas consultas no tienen que esperar en las colas detrás de otras consultas más largas. SQA solamente da prioridad a las consultas de corta ejecución y a las consultas que están en una cola definida por el usuario. Con SQA, las consultas cortas se ejecutan con mayor rapidez y los usuarios tardan menos en ver los resultados.

Si habilita SQA, podrá reducir las colas de administración de cargas de trabajo (WLM) dedicadas a la ejecución de consultas cortas. Además, las consultas de larga ejecución no necesitan competir con las consultas cortas por los slots de las colas, por lo que las colas de WLM se pueden configurar para que utilicen menos slots de consulta. Si se utiliza una simultaneidad más baja, el rendimiento de las consultas aumenta y el rendimiento de todo el sistema mejora con la mayoría de las cargas de trabajo.

Las instrucciones CREATE TABLE AS (CTAS) y las consultas de solo lectura, como las instrucciones SELECT, son aptas para SQA.

Amazon Redshift utiliza un algoritmo de machine learning para analizar cada una de las consultas que reúnan los requisitos necesarios y predecir su tiempo de ejecución. De forma predeterminada, WLM asigna dinámicamente un valor para el tiempo de ejecución máximo de SQA en función del análisis de la carga de trabajo del clúster. Como alternativa, puede especificar un valor fijo comprendido entre 1 y 20 segundos. Si el tiempo de ejecución previsto de la consulta es inferior al tiempo de ejecución máximo de SQA definido o asignado dinámicamente y la consulta está esperando en una cola de WLM, SQA separa la consulta de las colas de WLM y la programa para su ejecución prioritaria. Si una consulta tarda más en ejecutarse que el tiempo de ejecución máximo de SQA, WLM la mueve a la primera cola de WLM coincidente en función de las reglas de asignación de colas de WLM. Con el tiempo, las predicciones mejoran, ya que SQA va aprendiendo de los patrones de consulta.

De forma predeterminada, la SQA está habilitada en el grupo de parámetros predeterminados y para todos los grupos de parámetros nuevos. Para desactivar la SQA en la consola de Amazon Redshift, modifique la configuración de la WLM de un grupo de parámetros y desmarque Enable short query acceleration (Habilitar la aceleración de consultas cortas). Como práctica recomendada, es

conveniente utilizar un número de slots de consulta de WLM igual o inferior a 15 para mantener un nivel óptimo en el rendimiento general del sistema. Para obtener información sobre cómo modificar las configuraciones de la WLM, consulte Configuración de la administración de cargas de trabajo en la Guía de administración de Amazon Redshift.

## Tiempo máximo de ejecución de consultas cortas

Si se habilita SQA, WLM establece en dynamic el tiempo de ejecución máximo de las consultas cortas de forma predeterminada. Le recomendamos que mantenga este valor para definir el tiempo de ejecución máximo de SQA. Puede anular el ajuste predeterminado especificando un valor fijo comprendido entre 1 y 20 segundos.

En algunos casos, podría plantearse la posibilidad de utilizar otros valores para el tiempo de ejecución máximo de SQA y mejorar así el rendimiento del sistema. En estos casos, analice la carga de trabajo para determinar el tiempo de ejecución máximo de la mayoría de las consultas cortas. La consulta siguiente devuelve el tiempo máximo de ejecución de consultas en un percentil aproximado del 70.º.

```
select least(greatest(percentile_cont(0.7)
within group (order by total_exec_time / 1000000) + 2, 2), 20)
from stl_wlm_query
where userid >= 100
and final_state = 'Completed';
```

Una vez que haya identificado un tiempo máximo de ejecución que funcione bien con la carga de trabajo, no necesitará modificarlo salvo que la carga de trabajo cambie considerablemente.

## Monitorización de SQA

Para saber si SQA está habilitado, ejecute la consulta siguiente. Si la consulta devuelve una fila, SQA está habilitado.

```
select * from stv_wlm_service_class_config
where service_class = 14;
```

En la consulta siguiente, se muestra el número de consultas que pasaron por cada cola de consultas (clase de servicio). También se muestra el tiempo de ejecución medio, el número de consultas con un tiempo de espera establecido en el percentil 90 y el tiempo de espera medio. Las consultas de SQA utilizan la clase de servicio 14.

```
select final_state, service_class, count(*), avg(total_exec_time),
percentile_cont(0.9) within group (order by total_queue_time), avg(total_queue_time)
from stl_wlm_query where userid >= 100 group by 1,2 order by 2,1;
```

Para saber cuántas consultas ha seleccionado SQA y se han completado correctamente, ejecute la consulta siguiente.

```
select a.queue_start_time, a.total_exec_time, label, trim(querytxt)
from stl_wlm_query a, stl_query b
where a.query = b.query and a.service_class = 14 and a.final_state = 'Completed'
order by b.query desc limit 5;
```

Para buscar las consultas seleccionadas por SQA que han agotado el tiempo de espera, ejecute la consulta siguiente.

```
select a.queue_start_time, a.total_exec_time, label, trim(querytxt)
from stl_wlm_query a, stl_query b
where a.query = b.query and a.service_class = 14 and a.final_state = 'Evicted'
order by b.query desc limit 5;
```

Para obtener más información sobre las consultas descartadas y, de manera más general, las acciones basadas en reglas que se pueden realizar en las consultas, consulte Reglas de monitoreo de consultas de WLM.

# Reglas de asignación de colas de WLM

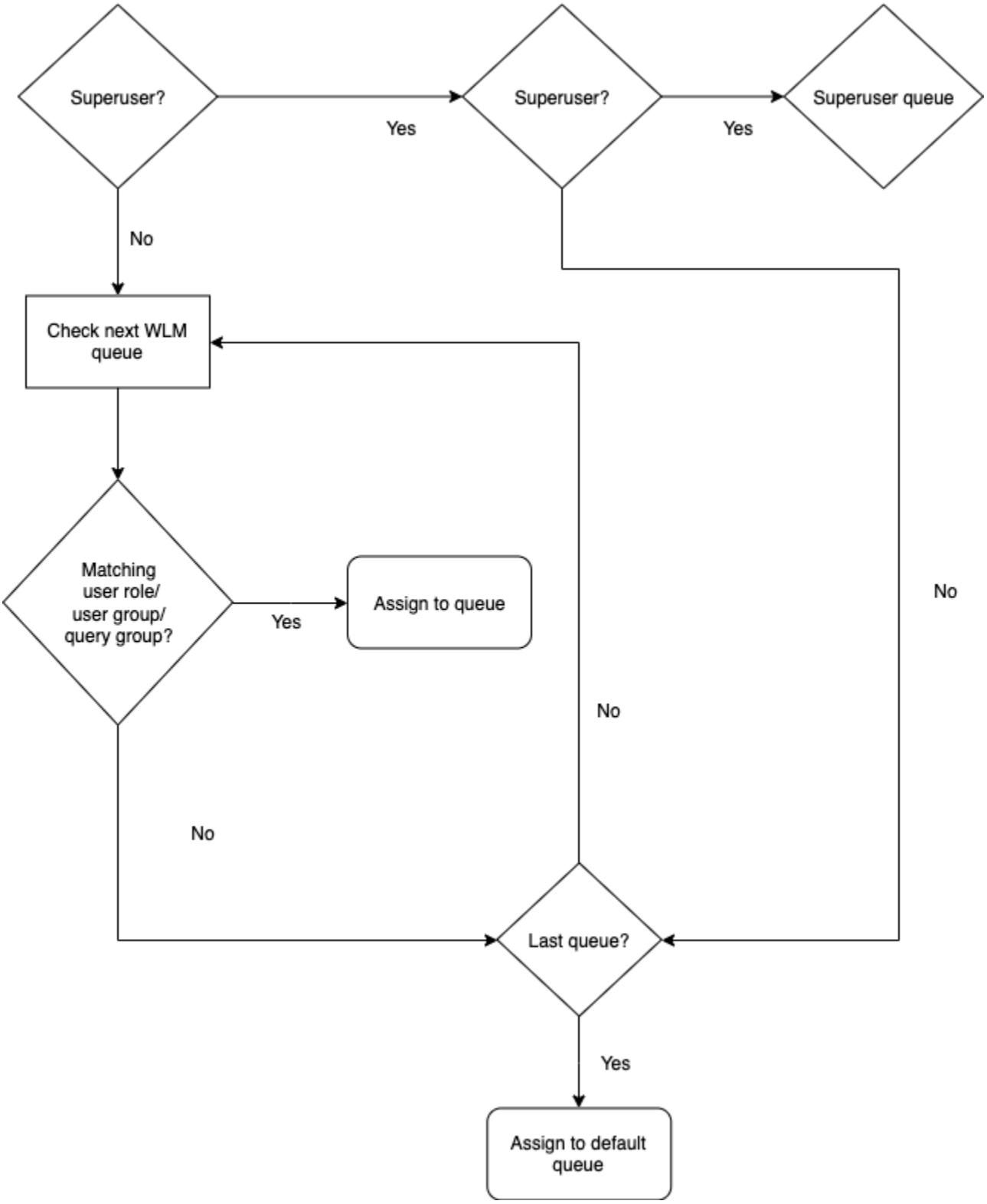
Con Amazon Redshift, puede controlar la asignación de recursos de memoria y CPU a las consultas de los usuarios mediante la definición de reglas de asignación de colas en una configuración de administración de carga de trabajo (WLM). En la siguiente sección, se describe la creación y la administración de reglas de asignación de colas de WLM para lograr una asignación de recursos eficiente y cumplir los acuerdos de nivel de servicio para diversas cargas de trabajo en Amazon Redshift.

Cuando un usuario ejecuta una consulta, WLM asigna la consulta a la primera cola coincidente, en función de las reglas de asignación de cola de WLM.

1. Si un usuario inició sesión como superusuario y ejecuta una consulta en el grupo de consultas con la etiqueta super usuario, la consulta se asigna a la cola superusuario.

2. Si un usuario es parte de un rol, pertenece a un grupo de usuarios mostrado o ejecuta una consulta dentro del grupo de consultas mostrado, la consulta se asigna a la primera cola coincidente.
3. Si una consulta no cumple con ningún criterio, la consulta se asigna a la cola predeterminada, que es la última cola definida en la configuración de WLM.

El siguiente diagrama ilustra cómo funcionan estas reglas.

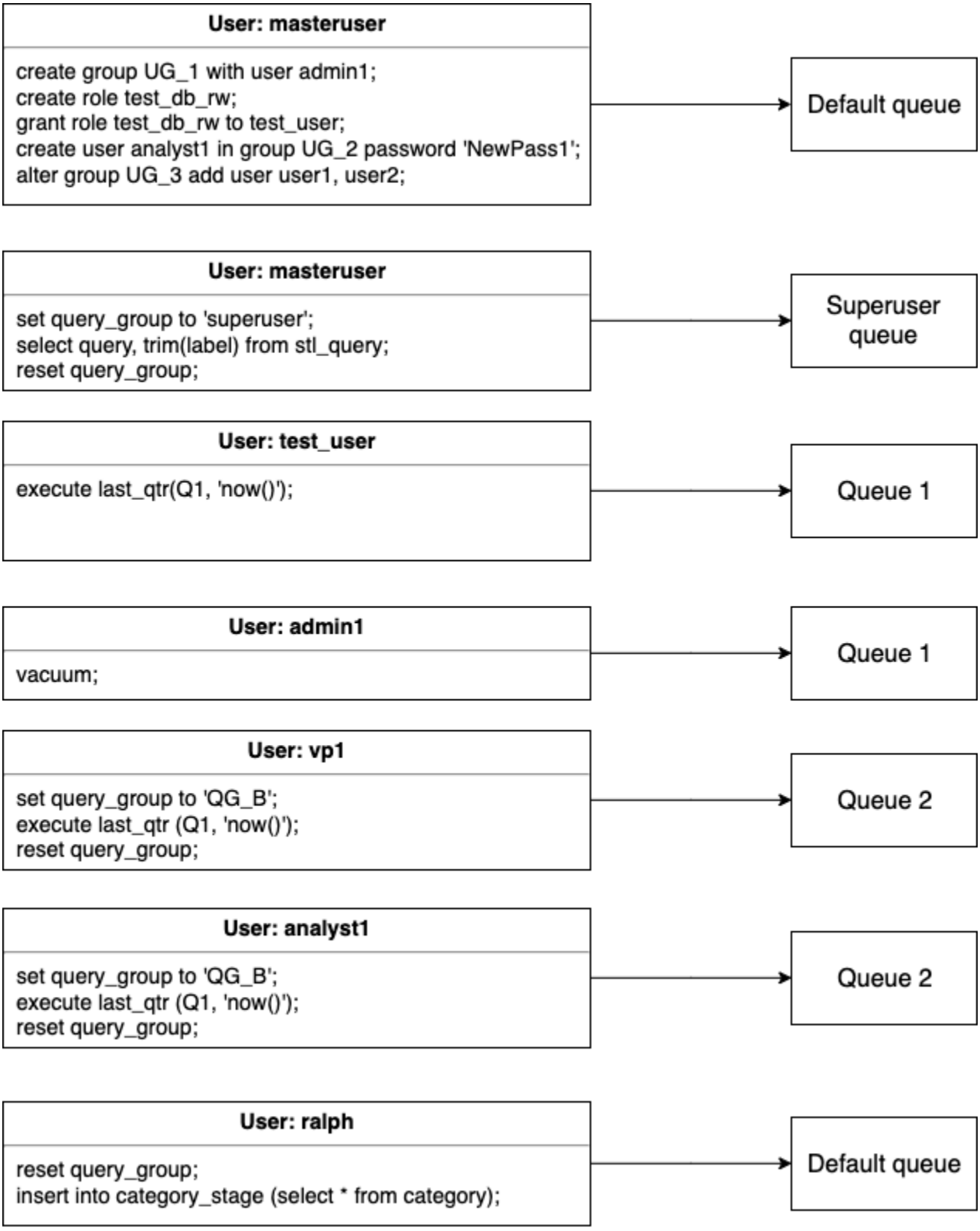

Superuser?
Yes
Superuser?
Yes
Superuser queue
No
Check next WLM queue
Matching user role/ user group/ query group?
Yes
Assign to queue
No
No
No
Last queue?
Yes
Assign to default queue

# Ejemplo de asignaciones de colas

En la siguiente tabla se muestra una configuración de WLM con la cola de superusuario y cuatro colas definidas por el usuario.

| Cola | Simultaneidad | Roles de usuario | User Groups (Grupos de usuarios) | Grupos de consultas |
|---|---|---|---|---|
| Superusuario | 1 | | | superusuario |
| 1 | 5 | test_db_rw | UG_1 | |
| 2 | 5 | | | QG_B |
| 3 | 5 | | UG_2 | QG_C |
| Predeterminado/a | 5 | | | |

La siguiente ilustración muestra el modo en que se asignan las consultas a las colas en la tabla anterior de acuerdo con los grupos de usuarios y de consultas. Para obtener información acerca del modo en que se asignan las consultas a los grupos de usuarios y de consultas en tiempo de ejecución, consulte Asignación de consultas a las colas más adelante en esta sección.




En este ejemplo, WLM realiza las siguientes asignaciones:

1. El primer conjunto de instrucciones muestra tres formas de asignar usuarios a los grupos de usuarios. Las instrucciones se ejecutan por el usuario `adminuser`, que no es miembro de un grupo de usuarios mencionado en ninguna cola de WLM. No se estableció ningún grupo de consultas, por lo que las instrucciones se dirigen a la cola predeterminada.
2. El usuario `adminuser` es un superusuario y el grupo de consultas se establece en `'superuser'`, por lo que la consulta se asigna a la cola de super usuario.
3. El usuario `test_user` se asigna al rol `test_db_rw` mostrado en la cola 1, por lo que la consulta se asigna a la cola 1.
4. El usuario `admin1` es miembro del grupo de usuarios mencionado en la cola 1, por lo que la consulta se asigna a la cola 1.
5. El usuario `vp1` no es miembro de ningún grupo de usuarios de la lista. El grupo de consultas se establece en `'QG_B'`, por lo que la consulta se asigna a la cola 2.
6. El usuario `analyst1` es miembro del grupo de usuarios mencionado en la cola 3, pero `'QG_B'` coincide con la cola 2, por lo que la consulta se asigna a la cola 2.
7. El usuario `ralph` no es miembro de ningún grupo de usuarios de la lista y el grupo de consultas se reinició, por lo que no hay colas coincidentes. La consulta se asigna a la cola predeterminada.

# Asignación de consultas a las colas

Con Amazon Redshift, puede administrar la simultaneidad de las cargas de trabajo y priorizar las consultas asignándolas a las colas. Las colas le permiten asignar recursos, como memoria y CPU, para diferentes tipos de consultas o usuarios, lo que garantiza que las consultas críticas se prioricen sobre las menos importantes. En las secciones siguientes, se describe cómo crear colas, configurar las propiedades y asignar consultas entrantes en función de los criterios que defina.

En los siguientes ejemplos se asignan consultas a las colas según los roles de usuarios, grupos de usuarios y grupos de consultas.

## Asignación de consultas a las colas en función de los roles de usuarios

Si se asigna un rol a un usuario y ese rol está adjunto a una cola, las consultas ejecutadas por ese usuario se asignan a esa cola. En el ejemplo siguiente se crea un rol de usuario denominado `sales_rw` y se asigna el usuario `test_user` a ese rol.

```
create role sales_rw;
```

```
grant role sales_rw to test_user;
```

También puede combinar los permisos de dos roles mediante la concesión explícita de un rol a otro. Al asignar un rol anidado a un usuario, se conceden permisos de ambos roles al usuario.

```
create role sales_rw;
create role sales_ro;
grant role sales_ro to role sales_rw;
grant role sales_rw to test_user;
```

Para ver la lista de usuarios a los que se les han asignado roles en el clúster, consulte la tabla SVV_USER_GRANTS. Para ver la lista de roles a los que se les han concedido roles en el clúster, consulte la tabla SVV_ROLE_GRANTS.

```
select * from svv_user_grants;
select * from svv_role_grants;
```

## Asignación de consultas a las colas en función de los grupos de usuarios

Si se muestra el nombre de un grupo de usuarios en la definición de una cola, las consultas que ejecuten los miembros de ese grupo de usuarios se asignan a la cola correspondiente. En el siguiente ejemplo, se crean grupos de usuarios y se añaden usuarios a los grupos mediante los comandos SQL CREAR USUARIO, CREATE GROUP y ALTER GROUP.

```
create group admin_group with user admin246, admin135, sec555;
create user vp1234 in group ad_hoc_group password 'vpPass1234';
alter group admin_group add user analyst44, analyst45, analyst46;
```

## Asignación de una consulta a un grupo de consultas

Puede asignar una consulta a una cola en tiempo de ejecución asignando la consulta al grupo de consultas adecuado. Utilice el comando SET para comenzar un grupo de consultas.

```
SET query_group TO group_label
```

Aquí, *group_label* es una etiqueta de grupo de consultas que aparece en la configuración de la WLM.

Todas las consultas que ejecute después del comando `SET query_group` se ejecutarán como miembros del grupo de consultas especificado hasta que restablezca el grupo de consultas o finalice su inició de sesión actual. Para obtener información acerca de la configuración y el restablecimiento de objetos de Amazon Redshift, consulte SET y RESET en la Referencia de comandos SQL.

Las etiquetas de grupos de consultas que especifique deben estar incluidas en la configuración de WLM actual o, de lo contrario, el comando SET query_group no tendrá efecto en las colas de consultas.

La etiqueta definida en la cláusula TO se captura en los registros de consultas de modo que pueda utilizarla para solucionar problemas. Para obtener información acerca de los parámetros de configuración query_group, consulte query_group en la referencia de configuración.

En el siguiente ejemplo, se ejecutan dos consultas como parte de la "prioridad" del grupo de consultas y luego se restablece el grupo de consultas.

```
set query_group to 'priority';
select count(*)from stv_blocklist;
select query, elapsed, substring from svl_qlog order by query desc limit 5;
reset query_group;
```

## Asignación de consultas a la cola de superusuario

Para asignar una consulta a la cola de superusuario, inicie sesión en Amazon Redshift como superusuario y, luego, ejecute la consulta en el grupo del superusuario. Cuando haya terminado, restablezca el grupo de consultas para que las consultas posteriores no se ejecuten en la cola de superusuario.

En el siguiente ejemplo, se asignan dos comandos para ejecutar en la cola de superusuario.

```
set query_group to 'superuser';

analyze;
vacuum;
reset query_group;
```

Consulte la tabla de catálogo del sistema PG_USER para ver una lista de superusuarios.

```
select * from pg_user where usesuper = 'true';
```

# Propiedades de configuración dinámicas y estáticas de WLM

Las propiedades de configuración de WLM son dinámicas o estáticas. Puede aplicar propiedades dinámicas a la base de datos sin reiniciar el clúster, pero las propiedades estáticas requieren el reinicio del clúster para que los cambios surtan efecto. Sin embargo, si cambia propiedades dinámicas y estáticas al mismo tiempo, debe reiniciar el clúster para que todos los cambios en las propiedades surtan efecto. Esto es así sin importar si las propiedades que se cambian son dinámicas o estáticas.

Aunque se apliquen propiedades dinámicas, el estado del clúster es `modifying`. El cambio entre una WLM automática y una manual es un cambio estático y requiere que un clúster se reinicie para surtir efecto.

La tabla siguiente indica qué propiedades de WLM son dinámicas o estáticas cuando se utiliza WLM automática o WLM manual.

| Propiedad de WLM | WLM automática | WLM manual |
|---|---|---|
| Grupos de consultas | Dinámico | Estático |
| Comodín de grupo de consultas | Dinámico | Estático |
| Grupos de usuarios | Dinámico | Estático |
| Comodín de grupo de usuarios | Dinámico | Estático |
| Roles de usuario | Dinámico | Estático |
| Comodín de rol de usuario | Dinámico | Estático |
| Simultaneidad en principal | No aplicable | Dinámico |
| Modo de escalado de simultaneidad | Dinámico | Dinámico |
| Habilitar aceleración de consultas cortas | No aplicable | Dinámico |

| Propiedad de WLM | WLM automática | WLM manual |
|---|---|---|
| Tiempo máximo de ejecución de consultas cortas | Dinámico | Dinámico |
| Porcentaje de memoria por utilizar | No aplicable | Dinámico |
| Tiempo de espera | No aplicable | Dinámico |
| Priority (Prioridad) | Dinámico | No aplicable |
| Agregar o eliminar colas | Dinámico | Estático |

Si añade una regla de monitorización de consultas (QMR), o modifica o elimina una QMR existente, el cambio se produce automáticamente sin necesidad de reiniciar el clúster.

> **Note**
>
> Cuando se utiliza WLM manual, si el valor de tiempo de espera se modifica, el nuevo valor se aplica a cualquier consulta que empiece a ejecutarse después de que cambie el valor. Si se modifican la simultaneidad o el porcentaje de la memoria que se utilizarán, Amazon Redshift pasa a la nueva configuración de forma dinámica. De manera que las consultas que se estén ejecutando actualmente no se vean afectadas por el cambio. Para obtener más información, consulte Asignación de memoria dinámica de WLM.

Temas


- Asignación de memoria dinámica de WLM
- Ejemplo de WLM dinámica


## Asignación de memoria dinámica de WLM

En cada cola, WLM crea un número de slots de consultas igual al nivel de simultaneidad de la cola. La cantidad de memoria asignada a un slot de consultas es igual al porcentaje de memoria asignado a la cola dividido por el número de slots. Si cambia la asignación de memoria o la simultaneidad, Amazon Redshift administra de manera dinámica la transición a la configuración de la WLM nueva.

Así, las consultas activas pueden ejecutar hasta su finalización mediante la cantidad de memoria asignada en este momento. Al mismo tiempo, Amazon Redshift se asegura de que el uso total de la memoria nunca supere el 100 % de memoria disponible.

El administrador de cargas de trabajo utiliza el siguiente proceso para administrar la transición:

1. WLM recalcula la asignación de memoria para cada slot de consultas nuevo.
2. Si un slot de consultas no está en uso de forma activa por una consulta en ejecución, WLM elimina el slot, lo que hace que la memoria esté disponible para slots nuevos.
3. Si un slot de consultas está en uso de forma activa, WLM espera a que se finalice la consulta.
4. A medida que las consultas activas se completan, se eliminan los slots vacíos y se libera la memoria asociada.
5. A medida que vuelve a estar disponible la memoria suficiente para añadir uno o más slots, se añaden slots nuevos.
6. Cuando todas las consultas que se estuvieron ejecutando en el momento de la modificación se completen, el número de slots iguala el nivel de simultaneidad nuevo y la transición a la nueva configuración de WLM se completa.

En efecto, las consultas que se están ejecutando cuando se produce el cambio siguen utilizando la asignación de memoria original. Las consultas que están en cola cuando el cambio tiene lugar, se dirigen a nuevos slots a medida que van estando disponibles.

Si las propiedades dinámicas de WLM se modifican durante el proceso de transición, WLM empieza la transición a la nueva configuración de inmediato, comenzando desde el estado actual. Para ver el estado de la transición, consulte la tabla de sistema STV_WLM_SERVICE_CLASS_CONFIG.

## Ejemplo de WLM dinámica

Con Amazon Redshift, puede administrar automáticamente la distribución de la carga de trabajo y la asignación de recursos en los clústeres de Amazon Redshift mediante WLM dinámica (Administración de la carga de trabajo). La WLM dinámica es un ejemplo de configuración de administración de la carga de trabajo (WLM) que ajusta dinámicamente las asignaciones de memoria en función de las demandas de carga de trabajo, lo que permite una simultaneidad y un rendimiento óptimos. En la siguiente sección, se proporcionan detalles sobre la implementación y la configuración de WLM dinámica para los clústeres de Amazon Redshift.

Suponga que el clúster de WLM está configurado con dos colas que utilizan las siguientes propiedades dinámicas.

| Cola | Simultaneidad | % de memoria por utilizar |
|---|---|---|
| 1 | 4 | 50% |
| 2 | 4 | 50% |

Ahora suponga que el clúster tiene 200 GB de memoria disponible para el procesamiento de consultas. (Este número es arbitrario y se usa únicamente a título ilustrativo). Como muestra la siguiente ecuación, a cada slot se le asignan 25 GB.

```
(200 GB * 50% ) / 4 slots  = 25 GB
```

A continuación, cambie el WLM para utilizar las siguientes propiedades dinámicas.

| Cola | Simultaneidad | % de memoria por utilizar |
|---|---|---|
| 1 | 3 | 75% |
| 2 | 4 | 25% |

Como muestra la siguiente ecuación, la nueva asignación de memoria para cada slot de la cola 1 es de 50 GB.

```
(200 GB * 75% ) / 3 slots = 50 GB
```

Suponga que las consultas A1, A2, A3 y A4 se ejecutan cuando se aplica la nueva configuración y las consultas B1, B2, B3 y B4 están en cola. WLM reconfigura de forma dinámica los slots de consultas del modo siguiente.

| Paso | Consultas en ejecución | Número de slots actual | Número de slots de destino | Memoria asignada | Memoria disponible |
|---|---|---|---|---|---|
| 1 | A1, A2, A3, A4 | 4 | 0 | 100 GB | 50 GB |
| 2 | A2, A3, A4 | 3 | 0 | 75 GB | 75 GB |
| 3 | A3, A4 | 2 | 0 | 50 GB | 100 GB |
| 4 | A3, A4, B1 | 2. | 1 | 100 GB | 50 GB |
| 5 | A4, B1 | 1 | 1 | 75 GB | 75 GB |
| 6 | A4, B1, B2 | 1 | 2 | 125 GB | 25 GB |
| 7 | B1, B2 | 0 | 2 | 100 GB | 50 GB |
| 8 | B1, B2, B3 | 0 | 3 | 150 GB | 0 GB |

1. WLM recalcula la asignación de memoria para cada slot de consultas. Originalmente, la cola 1 tenía asignados 100 GB. La cola nueva tiene una asignación total de 150 GB, por lo que la cola nueva tiene de inmediato 50 GB disponibles. La cola 1 ahora utiliza cuatro slots y el nivel de simultaneidad nuevo es de tres slots, por lo no se añaden slots nuevos.
2. Cuando una consulta finaliza, el slot se elimina y se liberan 25 GB. La cola 1 ahora tiene tres slots y 75 GB de memoria disponible. La nueva configuración requiere 50 GB para cada slot nuevo, pero el nivel de simultaneidad nuevo es de tres slots, por lo no se añaden slots nuevos.
3. Cuando una segunda consulta finaliza, el slot se elimina y se liberan 25 GB. La cola 1 ahora tiene dos slots y 100 GB de memoria libre.
4. Se añade un slot nuevo que utiliza 50 GB de la memoria libre. La cola 1 ahora tiene tres slots y 50 GB de memoria libre. Las consultas en cola ahora se pueden dirigir al slot nuevo.
5. Cuando una tercera consulta finaliza, el slot se elimina y se liberan 25 GB. La cola 1 ahora tiene dos slots y 75 GB de memoria libre.
6. Se añade un slot nuevo que utiliza 50 GB de la memoria libre. La cola 1 ahora tiene tres slots y 25 GB de memoria libre. Las consultas en cola ahora se pueden dirigir al slot nuevo.

7. Cuando la cuarta consulta finaliza, el slot se elimina y se liberan 25 GB. La cola 1 ahora tiene dos slots y 50 GB de memoria libre.
8. Se añade un slot nuevo que utiliza los 50 GB de memoria libre. La cola 1 ahora tiene tres slots con 50 GB cada uno y se ha asignado toda la memoria disponible.

Se completó la transición y todos los slots de consultas están disponibles para las consultas en cola.

# Reglas de monitoreo de consultas de WLM

En la administración de cargas de trabajo (WLM) de Amazon Redshift, las reglas de monitoreo de consultas definen los límites de rendimiento basados en métricas para las colas de WLM y especifican la acción que se debe realizar cuando una consulta supere esos límites. Por ejemplo, para una cola dedicada a consultas de ejecución corta, puede crear una regla que cancele las consultas que se ejecuten durante más de 60 segundos. Para hacer un seguimiento de las consultas mal diseñadas, puede disponer de otra regla que registre las consultas que contienen bucles anidados.

Puede definir las reglas de monitorización de consultas como parte de la configuración de la administración de cargas de trabajo (WLM). Puede definir hasta 25 reglas para cada cola, con un límite de 25 reglas para todas las colas. Cada regla incluye hasta tres condiciones, o predicados, y una acción. Un predicado se compone de una métrica, una condición de comparación (=, < o >) y un valor. Si se cumplen todos los predicados para cualquier regla, se desencadena la acción de esa regla. Las acciones de reglas posibles son log, hop y abort, como se explica a continuación.

Las reglas de una cola determinada aplican únicamente a las consultas que se ejecuten en esa cola. Una regla es independiente de otras reglas.

WLM evalúa las métricas cada 10 segundos. Amazon Redshift aplica reglas de monitorización de consultas en la consulta secundaria cuando las consultas se reescriben automáticamente. Si se activa más de una regla durante el mismo periodo, WLM elige la regla con la acción más grave. Si la acción de dos reglas tiene la misma gravedad, WLM ejecuta las reglas en orden alfabético, según el nombre de la regla. Si la acción es hop o abort, se registra la acción y la consulta se desaloja de la cola. Si la acción es log, la consulta continúa ejecutándose en la cola. WLM inicia solo una acción log por consulta, por regla. Si la cola contiene otras reglas, estas permanecerán en vigor. Si la acción es hop y la consulta se dirige a otra cola, las reglas aplican para la nueva cola. Para obtener más información sobre el monitoreo de consultas y el seguimiento de las acciones realizadas en consultas específicas, consulte la recopilación de ejemplos en Aceleración de consultas cortas.

Cuando se cumplen todos los predicados de una regla, WLM escribe una fila a la tabla de sistema STL_WLM_RULE_ACTION. Además, Amazon Redshift registra en STV_QUERY_METRICS las métricas de las consultas que están en ejecución actualmente. Las métricas de consultas completadas se almacenan en STL_QUERY_METRICS.

## Definición de una regla de monitoreo de consultas

Las reglas de monitoreo de consultas se crean durante la configuración de WLM, que se establece durante la definición del grupo de parámetros del clúster.

Puede crear reglas con la AWS Management Console o mediante programación con JSON.

> **Note**
>
> Si decide crear reglas mediante programación, recomendamos encarecidamente utilizar la consola para generar el JSON que incluirá en la definición del grupo de parámetros. Para obtener más información, consulte Creación o modificación de una regla de monitoreo de consultas mediante la consola y Configuración de los valores de parámetros mediante la AWS CLI en la Guía de administración de Amazon Redshift.

Para definir una regla de monitorización de consultas, especifique los elementos siguientes:

- Un nombre de regla: los nombres de reglas deben ser únicos dentro de la configuración de la WLM. Los nombres de reglas pueden tener un máximo de 32 caracteres alfanuméricos o guiones bajos y no pueden contener espacios ni comillas. Puede tener hasta 25 reglas por cola y el límite total para todas las colas es 25 reglas.
- Uno o más predicados: puede tener hasta tres predicados por regla. Si se cumplen todos los predicados para cualquier regla, se desencadena la acción asociada. Un predicado se define mediante un nombre de métrica, un operador (=, <, o >) y un valor. Un ejemplo es `query_cpu_time > 100000`. Para obtener una lista de métricas y ejemplos de valores para diferentes métricas, consulte Métricas de supervisión de consultas para Amazon Redshift aprovisionado más adelante en esta sección.
- Una acción: si se desencadena más de una regla, la WLM selecciona la regla con la acción más severa. Las acciones posibles, en orden ascendente de gravedad, son:
  - Registro: registra la información acerca de la consulta en la tabla de sistema STL_WLM_RULE_ACTION. Utilice la acción Log cuando solo desee escribir un registro. WLM

crea como máximo un registro por consulta, por regla. Después de llevar a cabo una acción log, las otras reglas permanecen vigentes y WLM continúa monitorizando la consulta.

- Salto (disponible solo con la WLM manual): registra la acción y transfiere la consulta a la siguiente cola coincidente. Si no existe otra cola coincidente, la consulta se cancela. QMR solo salta instrucciones CREATE TABLE AS (CTAS) y consultas de solo lectura, como instrucciones SELECT. Para obtener más información, consulte Salto de cola de consultas de WLM.
- Anulación: registra la acción y cancela la consulta. QMR no detiene las instrucciones COPY ni las operaciones de mantenimiento, como ALTER, ANALYZE y VACUUM.
- Modificación de prioridad (disponible solo con la WLM automática): cambia la prioridad de una consulta.

Para limitar el tiempo de ejecución de consultas, le recomendamos que cree una regla de monitorización de consultas en lugar de usar el tiempo de espera de WLM. Por ejemplo, puede establecer `max_execution_time` en 50.000 milisegundos como se muestra en el siguiente fragmento JSON.

```
"max_execution_time": 50000
```

Sin embargo, le recomendamos que defina una regla de supervisión de consultas equivalente. En el ejemplo siguiente se muestra una regla de supervisión de consultas que establece `query_execution_time` en 50 segundos:

```
"rules":
[
    {
        "rule_name": "rule_query_execution",
        "predicate": [
            {
                "metric_name": "query_execution_time",
                "operator": ">",
                "value": 50
            }
        ],
        "action": "abort"
    }
]
```

Si desea conocer los pasos necesarios para crear o modificar una regla de monitoreo de consultas, consulte Creación o modificación de una regla de monitoreo de consultas mediante la consola y Propiedades del parámetro wlm_json_configuration en la Guía de administración de Amazon Redshift.

Puede encontrar más información acerca de las reglas de monitorización de consultas en los temas siguientes:

- Métricas de supervisión de consultas para Amazon Redshift aprovisionado
- Plantillas de reglas de monitoreo de consultas
- Creación de una regla con la consola
- Configuración de la administración de cargas de trabajo
- Tablas y vistas de sistema para reglas de monitoreo de consultas

## Métricas de supervisión de consultas para Amazon Redshift aprovisionado

En la siguiente tabla se describen las métricas utilizadas en las reglas de monitorización de consultas. (Estas métricas son distintas de las métricas almacenadas en las tablas del sistema STV_QUERY_METRICS y STL_QUERY_METRICS).

Para una métrica determinada, se hace un seguimiento del umbral de rendimiento al nivel de la consulta o del segmento. Para obtener más información acerca de los segmentos y pasos, consulte Flujo de trabajo de planificación y ejecución de consultas.

> **Note**
>
> El parámetro Tiempo de espera de WLM es distinto de las reglas de monitorización de consultas.

| Métrica | Nombre | Descripción |
|---|---|---|
| Query CPU time | `query_cpu_time` | Tiempo de CPU, en segundos, que utiliza la consulta. `CPU time` es diferente de `Query execution time`.<br><br>Los valores válidos son de 0 a 999 999. |

| Métrica | Nombre | Descripción |
| --- | --- | --- |
| Blocks read | `query_blocks_read` | Número de bloques de datos de 1 MB leídos por la consulta.<br><br>Los valores válidos son de 0 a 1 048 575. |
| Scan row count | `scan_row_count` | La cantidad de filas en un paso del examen. La cantidad de filas es el número total de filas emitidas antes de filtrar las filas marcadas para eliminación (filas fantasma) y antes de aplicar los filtros de consulta definidos por el usuario.<br><br>Los valores válidos son de 0 a 999 999 999 999 999. |
| Query execution time | `query_execution_time` | Tiempo de ejecución transcurrido para una consulta, en segundos. El tiempo de ejecución no incluye el tiempo de espera empleado en una cola.<br><br>Los valores válidos son de 0 a 86 399. |
| Tiempo de cola de consulta | `query_queue_time` | Tiempo de espera en una cola, en segundos.<br><br>Los valores válidos son de 0 a 86 399. |
| Uso de la CPU | `query_cpu_usage_percent` | Porcentaje de capacidad de la CPU que utilizó la consulta.<br><br>Los valores válidos son de 0 a 6399. |
| Memory to disk | `query_temp_blocks_to_disk` | Espacio temporal en disco utilizado para escribir resultados intermedios, en bloques de 1 MB.<br><br>Los valores válidos son de 0 a 319 815 679. |

| Métrica | Nombre | Descripción |
| --- | --- | --- |
| CPU skew | `cpu_skew` | La proporción entre el uso máximo de la CPU para cualquier sector y el uso promedio de la CPU para todos los sectores. La métrica se define a nivel segmento.<br><br>Los valores válidos son de 0 a 99. |
| I/O skew | `io_skew` | La proporción entre la lectura máxima de bloques (E/S) para cualquier sector y la lectura promedio de bloques para todos los sectores. La métrica se define a nivel segmento.<br><br>Los valores válidos son de 0 a 99. |
| Rows joined | `join_row_count` | La cantidad de filas procesadas en un paso de combinación.<br><br>Los valores válidos son de 0 a 999 999 999 999 999. |
| Nested loop join row count | `nested_loop_join_row_count` | El número de filas en una combinación de bucle anidado.<br><br>Los valores válidos son de 0 a 999 999 999 999 999. |
| Return row count | `return_row_count` | La cantidad de filas devueltas por la consulta.<br><br>Los valores válidos son de 0 a 999 999 999 999 999. |
| Segment execution time | `segment_execution_time` | Tiempo de ejecución transcurrido para un segmento único, en segundos. Para evitar o reducir errores de muestreo, incluya `segment_execution_time > 10` en las reglas.<br><br>Los valores válidos son de 0 a 86 388. |

| Métrica | Nombre | Descripción |
| --- | --- | --- |
| Spectrum scan row count | `spectrum_scan_row_count` | El número de filas de datos en Amazon S3 analizadas por una consulta de Amazon Redshift Spectrum.<br><br>Los valores válidos son de 0 a 999 999 999 999 999. |
| Spectrum scan size | `spectrum_scan_size_mb` | La cantidad de datos en Amazon S3, en MB, analizados por una consulta de Amazon Redshift Spectrum.<br><br>Los valores válidos son de 0 a 999 999 999 999 999. |
| Prioridad de consulta | `query_priority` | La prioridad de la consulta.<br><br>Los valores válidos son `HIGHEST`, `HIGH`, `NORMAL`, `LOW` y `LOWEST`. Al comparar `query_priority` utilizando operadores mayor que (>) y menor que (<), `HIGHEST` es mayor que `HIGH`, `HIGH` es mayor que `NORMAL`, etc. |

**Note**

- La acción de salto no se admite con el predicado `query_queue_time`. Es decir, se omiten las reglas de salto definidas cuando se cumplen los requisitos del predicado `query_queue_time`.
- Los tiempos de ejecución de segmentos cortos pueden ocasionar errores de muestreo con otras métricas, como `io_skew` y `query_cpu_usage_percent`. Para evitar o reducir errores de muestreo, incluya el tiempo de ejecución de segmentos en las reglas. Un buen punto de partida es `segment_execution_time > 10`.

La vista SVL_QUERY_METRICS muestra las métricas de consultas completadas. La vista SVL_QUERY_METRICS_SUMMARY muestra los valores máximos de métricas de consultas completadas. Utilice los valores de estas vistas como ayuda para determinar los valores de umbral para definir las reglas de monitorización de consultas.

## Métricas de supervisión de consultas para Amazon Redshift sin servidor

En la siguiente tabla se describen las métricas utilizadas en las reglas de supervisión de consultas para Amazon Redshift sin servidor.

| Métrica | Nombre | Descripción |
|---|---|---|
| Blocks read | `max_query_blocks_read` | Número de bloques de datos de 1 MB leídos por la consulta.<br><br>Los valores válidos son de 0 a 1 048 575. |
| Scan row count | `max_scan_row_count` | La cantidad de filas en un paso del examen. La cantidad de filas es el número total de filas emitidas antes de filtrar las filas marcadas para eliminación (filas fantasma) y antes de aplicar los filtros de consulta definidos por el usuario.<br><br>Los valores válidos son de 0 a 999 999 999 999 999. |
| Query execution time | max_query_execution_time | Tiempo de ejecución transcurrido para una consulta, en segundos. El tiempo de ejecución no incluye el tiempo de espera empleado en una cola. Si una consulta supera el tiempo de ejecución establecido, Amazon Redshift sin servidor la detiene.<br><br>Los valores válidos son de 0 a 86 399. |
| Tiempo de cola de consulta | `max_query_queue_time` | Tiempo de espera en una cola, en segundos.<br><br>Los valores válidos son de 0 a 86 399. |

| Métrica | Nombre | Descripción |
| --- | --- | --- |
| Memory to disk | `max_query_temp_blocks_to_disk` | Espacio temporal en disco utilizado para escribir resultados intermedios, en bloques de 1 MB.<br><br>Los valores válidos son de 0 a 319 815 679. |
| Rows joined | `max_join_row_count` | La cantidad de filas procesadas en un paso de combinación.<br><br>Los valores válidos son de 0 a 999 999 999 999 999. |
| Nested loop join row count | `max_nested_loop_join_row_count` | El número de filas en una combinación de bucle anidado.<br><br>Los valores válidos son de 0 a 999 999 999 999 999. |

> **Note**
>
> - La acción de salto no se admite con el predicado `max_query_queue_time`. Es decir, se omiten las reglas de salto definidas cuando se cumplen los requisitos del predicado `max_query_queue_time`.
> - Los tiempos de ejecución de segmentos cortos pueden ocasionar errores de muestreo con otras métricas, como `max_io_skew` y `max_query_cpu_usage_percent`.

## Plantillas de reglas de monitoreo de consultas

Cuando agrega una regla utilizando la consola de Amazon Redshift, puede elegir crear una regla a partir de una plantilla predefinida. Amazon Redshift crea una nueva regla con un conjunto de predicados y completa los predicados con valores predeterminados. La acción predeterminada es log. Puede modificar los predicados y las acciones para que cumplan con el caso de uso.

En la tabla siguiente se muestran las plantillas disponibles.

| Nombre de la plantilla | Predicados | Descripción |
| --- | --- | --- |
| Combinación de bucle anidado | `nested_loop_join_row_count > 100` | Una combinación de bucles anidados puede indicar un predicado de combinación incompleto, que suele ocasionar un conjunto de retorno muy grande (un producto cartesiano). Utilice un número de filas bajo para encontrar una consulta potencialmente descontrolada con rapidez. |
| Query returns a high number of rows | `return_row_count > 1000000` | Si dedica una cola a consultas de ejecución corta y sencilla, puede incluir una regla que encuentre consultas que devuelvan un número de filas alto. La plantilla utiliza un valor predeterminado de 1 millón de filas. Para algunos sistemas, puede considerar que un millón de filas es un número alto, o en sistemas más grandes, mil millones de consultas o más puede ser alto. |
| Join with a high number of rows | `join_row_count > 1000000000` | Un paso de combinación que incluya un número de filas inusualmente alto puede indicar la necesidad de más filtros restrictivos. La plantilla utiliza un valor predeterminado de mil millones de filas. Para una cola personalizada (de una única vez) que está destinada a consultas sencillas y rápidas, puede utilizar un número menor. |
| High disk usage when writing intermediate results | `query_temp_blocks_to_disk > 100000` | Cuando las consultas en ejecución utilizan más RAM de sistema que la disponible, el motor de ejecución de consultas escribe resultados intermedios en el disco (memoria vertida). Por lo general, esta condición es el resultado de una consulta inadecuada, que normalmente también es la consulta que utiliza el mayor espacio en disco. El umbral aceptable de uso |

| Nombre de la plantilla | Predicados | Descripción |
| --- | --- | --- |
| | | de disco varía en función del tipo y el número de nodos del clúster. La plantilla utiliza un valor predeterminado de 100 000 bloques o 100 GB. Para un clúster pequeño, puede utilizar un número menor. |
| Long running query with high I/O skew | `segment_execution_time > 120` y `io_skew > 1.30` | El sesgo de E/S se produce cuando un sector del nodo tiene una tasa de E/S mucho mayor que los otros sectores. Como punto de partida, un sesgo de 1,30 (1,3 veces el promedio) es considerado alto. El sesgo de E/S alto no siempre es un problema, pero cuando se combina con un tiempo de ejecución de consulta prolongado, puede indicar un problema con el estilo de distribución o la clave de ordenación. |

## Tablas y vistas de sistema para reglas de monitoreo de consultas

Cuando se cumplen todos los predicados de una regla, WLM escribe una fila a la tabla de sistema STL_WLM_RULE_ACTION. Esta fila contiene detalles para la consulta que activaron la regla y la acción resultante.

Además, Amazon Redshift registra métricas de consultas para las siguientes tablas de sistema y vistas.

- La tablaSTV_QUERY_METRICS muestra las métricas de consultas en ejecución.
- La tabla STL_QUERY_METRICS registra las métricas de consultas completadas.
- La vista SVL_QUERY_METRICS muestra las métricas de consultas completadas.
- La vista SVL_QUERY_METRICS_SUMMARY muestra los valores máximos de métricas de consultas completadas.

# Tablas y vistas de sistema de WLM

La WLM configura las colas de consultas según las clases de servicio de la WLM, las cuales se definen internamente. Amazon Redshift crea varias colas internas en función de estas clases de servicio junto con las colas definidas en la configuración de la WLM. Los términos cola y clase de servicios suelen utilizarse indistintamente en las tablas de sistema. La cola de superusuario utiliza la clase de servicio 5. Las colas definidas por el usuario utilizan la clase de servicio 6 o mayores.

Puede ver el estado de consultas, colas y clases de servicios mediante las tablas de sistema específicas de WLM. Consulte las siguientes tablas de sistema para:

- Ver a qué consultas se les hace un seguimiento y qué recursos son asignados por el administrador de cargas de trabajo.
- Ver a qué cola se asignó una consulta.
- Ver el estado de una consulta a la que el administrador de cargas de trabajo le hace un seguimiento.

| Nombre de la tabla | Descripción |
|---|---|
| STL_WLM_ERROR | Contiene un registro de los eventos de error relacionados a WLM. |
| STL_WLM_QUERY | Enumera las consultas a las que WLM hace un seguimiento. |
| STV_WLM_CLASSIFICATION_CONFIG | Muestra las reglas de clasificación actuales para WLM. |
| STV_WLM_QUERY_QUEUE_STATE | Registra el estado actual de las colas de consultas. |
| STV_WLM_QUERY_STATE | Proporciona una instantánea del estado actual de las consultas a las que WLM hace un seguimiento. |
| STV_WLM_QUERY_TASK_STATE | Contiene el estado actual de las tareas de consultas. |
| STV_WLM_SERVICE_CLASS_CONFIG | Registra las configuraciones de las clases de servicios para WLM. |

| Nombre de la tabla | Descripción |
|---|---|
| STV_WLM_SERVICE_CLASS_STATE | Tiene el estado actual de las clases de servicios. |
| STL_WLM_RULE_ACTION | Registra detalles de las acciones que surgen de las reglas de supervisión de consultas de la Workload Management (WLM, Administración de cargas de trabajo) asociadas a colas definidas por el usuario. |
| STV_WLM_QMR_CONFIG | Registra la configuración para las Query Monitoring Rules (QMR, Reglas de monitorización de consultas) de WLM. |

Utilice el ID de tarea para hacer un seguimiento de una consulta en las tablas de sistema. En el siguiente ejemplo, se muestra cómo obtener el ID de tarea de la consulta de usuario que se envió más recientemente:

```
select task from stl_wlm_query where exec_start_time =(select max(exec_start_time) from
 stl_wlm_query);

task
------
137
(1 row)
```

En el siguiente ejemplo, se muestran las consultas que actualmente se están ejecutando o que están esperando en varias clases de servicios (colas). Esta consulta es útil para hacer un seguimiento de la carga de trabajo general simultánea para Amazon Redshift:

```
select * from stv_wlm_query_state order by query;


xid |task|query|service_| wlm_start_  |  state  |queue_ | exec_
    |    |     |class   | time        |         |time   | time
----+----+-----+--------+-------------+---------+-------+--------
2645| 84 | 98  | 3      | 2010-10-... |Returning|   0   | 3438369
2650| 85 | 100 | 3      | 2010-10-... |Waiting  |   0   | 1645879
2660| 87 | 101 | 2      | 2010-10-... |Executing|   0   | 916046
2661| 88 | 102 | 1      | 2010-10-... |Executing|   0   | 13291
```

```
(4 rows)
```

## ID de la clase de servicio WLM

La siguiente tabla enumera los ID asignados a las clases de servicio.

| ID | Clase de servicio |
| --- | --- |
| 1–4 | Reservado para el uso del sistema. |
| 5 | Utilizado por la cola de superusuario. |
| 6–13 | Utilizado por las colas de WLM manual definidas en la configuración de WLM. |
| 14 | Utilizado por la aceleración de consultas cortas. |
| 15 | Reservado para actividades de mantenimiento ejecutadas por Amazon Redshift. |
| 100–107 | Utilizado por la cola de WLM automática cuando auto_wlm es verdadero. |

# Seguridad de base de datos

La seguridad de la base de datos se administra al controlar los usuarios que pueden obtener acceso a determinados objetos de la base de datos. A los usuarios se les pueden asignar roles o grupos, y los permisos que concede a los usuarios, roles o grupos deciden a qué objetos de la base de datos pueden acceder.

Temas


- Información general de la seguridad de Amazon Redshift
- Permisos de usuario de base de datos predeterminados
- Superusuarios
- Usuarios
- Grupos
- Esquemas
- Control de acceso basado en roles (RBAC)
- Seguridad de nivel básico
- Seguridad de los metadatos
- Enmascaramiento de datos dinámico
- Permisos acotados


El acceso a los objetos de la base de datos depende de los permisos que concede a usuarios o roles. Las siguientes directrices resumen cómo funciona la seguridad de las bases de datos:

- De manera predeterminada, solo se conceden permisos al propietario del objeto.
- Los usuarios de base de datos de Amazon Redshift son usuarios designados que pueden conectarse a una base de datos. Un usuario recibe permisos de dos formas: explícitamente, si esos permisos se le asignan directamente a la cuenta, o implícitamente, si pertenece a un grupo que tiene permisos concedidos.
- Los grupos son conjuntos de usuarios a los que se les pueden asignar privilegios de forma colectiva con objeto de agilizar el mantenimiento de la seguridad.
- Los esquemas son colecciones de tablas y otros objetos de la base de datos. Los esquemas son similares a los directorios del sistema de archivos, excepto que los esquemas no se pueden anidar. Los usuarios pueden recibir acceso a un único esquema o a varios.

Además, Amazon Redshift emplea las siguientes características para ofrecerle un control más preciso sobre los usuarios que tienen acceso a los objetos de la base de datos:

- El control de acceso basado en roles (RBAC) le permite asignar permisos a los roles que puede aplicar a los usuarios, lo que le permite controlar los permisos en grupos de usuarios grandes. A diferencia de los grupos, los roles pueden heredar permisos de otros roles.

  La seguridad a nivel de fila (RLS) le permite definir políticas que restringen el acceso a las filas que elija y, después, aplicarlas a usuarios o grupos.

  El enmascaramiento dinámico de datos (DDM) protege aún más los datos al transformarlos en el tiempo de ejecución de la consulta para que los usuarios puedan acceder a ellos sin exponer detalles confidenciales.

Para ver ejemplos de implementaciones de seguridad, consulte Ejemplo del control de acceso de usuarios y grupos.

Para obtener más información acerca de la protección de los datos, consulte Seguridad en Amazon Redshift en la Guía de administración de Amazon Redshift.

# Información general de la seguridad de Amazon Redshift

La seguridad de base de datos de Amazon Redshift es distinta a otros tipos de seguridad de Amazon Redshift. Además de la seguridad de base de datos, que se describe en esta sección, Amazon Redshift proporciona estas características para administrar la seguridad:

- Credenciales de inicio de sesión: el acceso a la consola de administración de AWS de Amazon Redshift se controla mediante los permisos de la cuenta de AWS. Para obtener más información, consulte Credenciales de inicio de sesión.
- Administración de acceso: para controlar el acceso a recursos de Amazon Redshift específicos, debe definir las cuentas de AWS Identity and Access Management (IAM). Para obtener más información, consulte Control de acceso a recursos de Amazon Redshift.
- Grupos de seguridad de los clústeres: para conceder a otros usuarios acceso de entrada a un clúster de Amazon Redshift, debe definir un grupo de seguridad de clúster y asociarlo con un clúster. Para obtener más información, consulte Grupos de seguridad de clústeres de Amazon Redshift.

- VPC: para proteger el acceso al clúster mediante un entorno de redes virtuales, puede lanzar el clúster en una Amazon Virtual Private Cloud (VPC). Para obtener más información, consulte Administración de clústeres en una Virtual Private Cloud (VPC).
- Cifrado del clúster: para cifrar los datos de todas las tablas creadas por el usuario, puede activar el cifrado del clúster cuando lance el clúster. Para obtener más información, consulte Clústeres de Amazon Redshift.
- Conexiones SSL: para cifrar la conexión entre el cliente SQL y el clúster, puede utilizar el cifrado de capa de conexión segura (SSL). Para obtener más información, consulte Conexión a un clúster con SSL.
- Cifrado de datos de carga: para cifrar los archivos de datos de carga de las tablas al cargarlos en Amazon S3, puede usar el cifrado del lado del servidor o el cifrado del lado del cliente. Cuando realiza cargas a partir de datos cifrados del lado del servidor, Amazon S3 se ocupa del descifrado de manera transparente. Cuando realiza cargas a partir de datos cifrados del lado del cliente, el comando COPY de Amazon Redshift descifra los datos a medida que se cargan en la tabla. Para obtener más información, consulte Carga de datos cifrados en Amazon S3.
- Datos en tránsito: para proteger los datos en tránsito en la nube de AWS, Amazon Redshift utiliza SSL con aceleración por hardware para comunicarse con Amazon S3, o Amazon DynamoDB en las operaciones COPY, UNLOAD, de copia de seguridad y de restauración.
- Control de acceso de nivel de columna: para tener un control de acceso de nivel de columna para los datos en Amazon Redshift, utilice instrucciones de concesión y revocación de nivel de columna sin tener que implementar un control de acceso basado en vistas ni emplear otro sistema.
- Control de seguridad de nivel de fila: para disponer de un control de seguridad de nivel de fila para los datos en Amazon Redshift, cree y asocie políticas a roles o usuarios que restrinjan el acceso a las filas definidas en la política.

# Permisos de usuario de base de datos predeterminados

Cuando crea un objeto de base de datos, usted es su propietario. De manera predeterminada, solo un superusuario o el propietario de un objeto pueden consultar, modificar o conceder permisos para el objeto. Para los usuarios que utilicen un objeto, debe conceder los permisos necesarios al usuario o grupo que contenga al usuario. Los superusuarios de bases de datos tienen los mismos permisos que los propietarios de bases de datos.

Amazon Redshift admite los siguientes permisos: SELECT, INSERT, UPDATE, DELETE, REFERENCES, CREATE, TEMPORARY y USAGE. Se asocian diferentes permisos a diferentes

tipos de objetos. Para obtener información acerca de los permisos de objetos de base de datos admitidos por Amazon Redshift, consulte el comando GRANT.

Solo el propietario tiene permiso para modificar o destruir un objeto.

De manera predeterminada, todos los usuarios tienen privilegios CREATE y USAGE para el esquema PUBLIC de una base de datos. Para dejar de permitir a los usuarios crear objetos en el esquema PUBLIC de una base de datos, utilice el comando REVOKE para eliminar ese permiso.

Para revocar un permiso que se haya concedido previamente, utilice el comando REVOKE. Los permisos del propietario del objeto, tales como los permisos DROP, GRANT y REVOKE, son implícitos y no se pueden conceder ni revocar. Los propietarios de objetos pueden revocar sus propios permisos ordinarios; por ejemplo, para hacer que una tabla sea de solo lectura para ellos mismos y para otras personas. Los superusuarios conservan todos los permisos, con independencia de los comandos GRANT y REVOKE.

## Superusuarios

Los superusuarios de bases de datos tienen los mismos permisos que los propietarios de bases de datos para todas las bases de datos.

El usuario admin (administrador), que es el usuario que creó cuando lanzó el clúster, es un superusuario.

Debe ser un superusuario para crear un superusuario.

Las tablas y vistas de sistema de Amazon Redshift están visibles solo para superusuarios o están visibles para todos los usuarios. Solo los superusuarios pueden enviar consultas a las tablas y vistas de sistema que son “visibles a los superusuarios”. Para obtener más información, consulta Vistas de monitoreo de SYS.

Los superusuarios pueden ver todas las tablas del catálogo. Para obtener más información, consulta Tablas de catálogos de sistema.

Un superusuario de base de datos omite todas las comprobaciones de permisos. Los superusuarios conservan todos los permisos, con independencia de los comandos GRANT y REVOKE. Tenga cuidado cuando utilice un rol de superusuario. Le recomendamos que realice la mayor parte de su trabajo con un rol que no sea el de un superusuario. Puede crear un rol de administrador con permisos más restrictivos. Para obtener más información acerca de la creación de roles, consulte Control de acceso basado en roles (RBAC).

Para crear un nuevo superusuario de base de datos, inicie sesión en la base de datos como superusuario y emita un comando CREATE USER o ALTER USER con el permiso CREATEUSER.

```
CREATE USER adminuser CREATEUSER PASSWORD '1234Admin';
ALTER USER adminuser CREATEUSER;
```

Para crear, modificar o eliminar un superusuario, utilice los mismos comandos para administrar usuarios. Para obtener más información, consulte Creación, modificación y eliminación de usuarios.

# Usuarios

Puede crear y administrar usuarios de bases de datos mediante los comandos SQL de Amazon Redshift CREATE USER y ALTER USER. También puede configurar el cliente SQL con controladores JDBC u ODBC de Amazon Redshift personalizados. Estos se encargan de administrar el proceso de creación de usuarios de bases de datos y contraseñas temporales como parte del proceso de inicio de sesión en las bases de datos.

Los controladores autentican a los usuarios de la base de datos en función de la autenticación de AWS Identity and Access Management (IAM). Si ya administra las identidades de los usuarios fuera de AWS, puede utilizar un proveedor de identidad (IdP) conforme con SAML 2.0 para administrar el acceso a los recursos de Amazon Redshift. Puede utilizar un rol de IAM para configurar el proveedor de identidad y AWS para permitir que los usuarios federados generen credenciales de base de datos temporales e inicien sesión en bases de datos de Amazon Redshift. Para obtener más información, consulte Uso de la autenticación de IAM para generar credenciales de usuario de base de datos.

Solo un superusuario de base de datos puede crear y eliminar usuarios de Amazon Redshift. Los usuarios se autentican cuando inician sesión en Amazon Redshift. Pueden ser propietarios de bases de datos y de objetos de bases de datos (por ejemplo, tablas). También pueden conceder permisos para esos objetos a usuarios, grupos y esquemas para controlar quién puede acceder a cada objeto. Los usuarios con derechos CREATE DATABASE pueden crear bases de datos y conceder permisos a esas bases de datos. Los superusuarios tienen permisos de propiedad de base de datos para todas las bases de datos.

## Creación, modificación y eliminación de usuarios

Los usuarios de base de datos se aplican globalmente en un clúster de almacenamiento de datos (no por base de datos individual).

- Para crear un usuario, utilice el comando CREAR USUARIO.
- Para crear un superusuario, utilice el comando CREAR USUARIO con la opción CREATEUSER.
- Para eliminar un usuario existente, use el comando DROP USER.
- Para modificar un usuario, por ejemplo, con el fin de cambiar una contraseña, utilice el comando ALTER USER.
- Para ver una lista de usuarios, realice una consulta de la tabla de catálogo PG_USER.

```
select * from pg_user;

  usename   | usesysid | usecreatedb | usesuper | usecatupd |  passwd  | valuntil |
 useconfig
------------+----------+-------------+----------+-----------+----------+----------
+-----------
 rdsdb      |        1 | t           | t        | t         | ******** |          |
 masteruser |      100 | t           | t        | f         | ******** |          |
 dwuser     |      101 | f           | f        | f         | ******** |          |
 simpleuser |      102 | f           | f        | f         | ******** |          |
 poweruser  |      103 | f           | t        | f         | ******** |          |
 dbuser     |      104 | t           | f        | f         | ******** |          |
(6 rows)
```

## Grupos

Los grupos son conjuntos de usuarios que han recibido todos ellos los permisos correspondientes asociados al grupo. Puede utilizar grupos para asignar permisos. Por ejemplo, puede crear diferentes grupos para ventas, administración y asistencia, y concederles a los usuarios de cada grupo el acceso adecuado a los datos que necesitan para su trabajo. Puede conceder o revocar permisos en el nivel del grupo, y esos cambios se aplicarán a todos los miembros del grupo, excepto a los superusuarios.

Consulte la tabla de catálogo del sistema PG_GROUP para ver una lista de todos los grupos de usuarios:

```
select * from pg_group;
```

Por ejemplo, para mostrar todos los usuarios de bases de datos por grupo, ejecute el siguiente SQL.

```
SELECT u.usesysid
```

```
,g.groname
,u.usename
FROM pg_user u
LEFT JOIN pg_group g ON u.usesysid = ANY (g.grolist)
```

## Creación, modificación y eliminación de grupos

Solo un superusuario puede crear, modificar o eliminar grupos.

Puede llevar a cabo las siguientes acciones:

- Para crear un grupo, use el comando CREATE GROUP.
- Para agregar o eliminar usuarios de un grupo existente, use el comando ALTER GROUP.
- Para eliminar un grupo, use el comando DROP GROUP. Este comando solo elimina el grupo, no los usuarios miembros.

## Ejemplo del control de acceso de usuarios y grupos

Este ejemplo crea grupos y usuarios y, a continuación, les concede diversos permisos para una base de datos de Amazon Redshift que se conecta a un cliente de aplicación web. Este ejemplo supone tres grupos de usuarios: usuarios normales de una aplicación web, usuarios avanzados de una aplicación web y desarrolladores web.

Para obtener información sobre cómo eliminar a un usuario de un grupo, consulte ALTER GROUP.

1. Cree los grupos donde se asignarán los usuarios. El siguiente conjunto de comandos crea tres grupos de usuarios diferentes:

   ```
   create group webappusers;

   create group webpowerusers;

   create group webdevusers;
   ```

2. Cree varios usuarios de base de datos con diferentes permisos y agréguelos a los grupos.

   a. Cree dos usuarios y añádalos al grupo WEBAPPUSERS:

   ```
   create user webappuser1 password 'webAppuser1pass'
   in group webappusers;
   ```

```
create user webappuser2 password 'webAppuser2pass'
in group webappusers;
```

b. Cree un usuario desarrollador web y agréguelo al grupo WEBDEVUSERS:

```
create user webdevuser1 password 'webDevuser2pass'
in group webdevusers;
```

c. Cree un superusuario. Este usuario tendrá derechos administrativos para crear otros usuarios:

```
create user webappadmin  password 'webAppadminpass1'
createuser;
```

3. Cree un esquema para asociarlo con las tablas de la base de datos que la aplicación web utiliza y conceda a diferentes grupos de usuarios acceso a este esquema:

a. Cree el esquema WEBAPP:

```
create schema webapp;
```

b. Conceda permisos USAGE al grupo WEBAPPUSERS:

```
grant usage on schema webapp to group webappusers;
```

c. Conceda permisos USAGE al grupo WEBPOWERUSERS:

```
grant usage on schema webapp to group webpowerusers;
```

d. Conceda permisos ALL al grupo WEBDEVUSERS:

```
grant all on schema webapp to group webdevusers;
```

Los usuarios y grupos básicos ya están configurados. Ahora puede modificar usuarios y grupos.

4. Por ejemplo, el siguiente comando modifica el parámetro search_path de WEBAPPUSER1.

```
alter user webappuser1 set search_path to webapp, public;
```

El comando SEARCH_PATH especifica el orden de búsqueda de esquemas para objetos de base de datos, como tablas y funciones, cuando un nombre simple que no tiene un esquema especificado hace referencia al objeto.

5. También puede añadir usuarios a un grupo después de crearlo; por ejemplo, añadir WEBAPPUSER2 al grupo WEBPOWERUSERS:

```
alter group webpowerusers add user webappuser2;
```

# Esquemas

Una base de datos contiene uno o más esquemas con nombre. Cada esquema de una base de datos contiene tablas u otros tipos de objetos con nombre. De manera predeterminada, una base de datos tiene un único esquema, que se denomina PUBLIC. Puede usar esquemas para agrupar objetos de la base de datos bajo un nombre común. Los esquemas son similares a los directorios del sistema de archivos, excepto que los esquemas no se pueden anidar.

Se pueden usar nombres de objetos de base de datos idénticos en diferentes esquemas de la misma base de datos sin generar conflictos. Por ejemplo, MY_SCHEMA y YOUR_SCHEMA pueden contener una tabla denominada MYTABLE. Los usuarios con los permisos necesarios pueden acceder a objetos de varios esquemas de una base de datos.

De manera predeterminada, se crea un objeto dentro del primer esquema en la ruta de búsqueda de la base de datos. Para obtener más información, consulte Ruta de búsqueda más adelante en esta sección.

Los esquemas pueden ser de ayuda en los problemas de organización y simultaneidad en un entorno multiusuario de las siguientes maneras:

- Para permitir que varios desarrolladores trabajen en la misma base de datos sin interferirse.
- Organizan objetos de base de datos en grupos lógicos para poder administrarlos más fácilmente.
- Les otorgan a las aplicaciones la capacidad de colocar sus objetos en esquemas separados para que sus nombres no colisionen con los nombres de los objetos utilizados por otras aplicaciones.

## Ruta de búsqueda

La ruta de búsqueda se define en el parámetro search_path con una lista de nombres de esquemas separados por comas. La ruta de búsqueda especifica el orden en el que se buscan los esquemas cuando un nombre simple que no incluye un calificador de esquema hace referencia a un objeto, como una tabla o función.

Si se crea un objeto sin especificar un esquema de destino, el objeto se agrega al primer esquema que aparece en la ruta de búsqueda. Cuando hay objetos con nombres idénticos en diferentes esquemas, un nombre de objeto que no especifica un esquema hará referencia al primer esquema de la ruta de búsqueda que contenga un objeto con ese nombre.

Para cambiar el esquema predeterminado para la sesión actual, use el comando SET.

Para obtener más información, consulte la descripción de search_path en la Referencia de configuración.

## Creación, modificación y eliminación de esquemas

Cualquier usuario puede crear o eliminar esquemas de los cuales es propietario.

Puede llevar a cabo las siguientes acciones:

- Para crear un esquema, use el comando CREATE SCHEMA.
- Para cambiar el propietario de un esquema, use el comando ALTER SCHEMA.
- Para eliminar un esquema y sus objetos, use el comando DROP SCHEMA.
- Para crear una tabla dentro de un esquema, cree la tabla con el formato nombre_de_esquema.nombre_de_tabla.

Consulte la tabla de catálogo del sistema PG_NAMESPACE para ver una lista de todos los esquemas:

```
select * from pg_namespace;
```

Consulte la tabla de catálogo del sistema PG_TABLE_DEF para ver una lista de las tablas que pertenecen a un esquema. Por ejemplo, la siguiente consulta devuelve una lista de tablas del esquema PG_CATALOG.

```
select distinct(tablename) from pg_table_def
where schemaname = 'pg_catalog';
```

## Permisos basados en esquemas

Los permisos basados en esquemas son determinados por el propietario del esquema:

- De manera predeterminada, todos los usuarios tienen privilegios CREATE y USAGE para el esquema PUBLIC de una base de datos. Para dejar de permitir a los usuarios crear objetos en el esquema PUBLIC de una base de datos, utilice el comando REVOKE con objeto de eliminar ese permiso.
- A menos que el propietario del objeto les conceda el permiso USAGE, los usuarios no pueden acceder a ninguno de los objetos de los esquemas de los que no son propietarios.
- Si los usuarios han recibido el permiso CREATE para un esquema que ha creado otro usuario, esos usuarios pueden crear objetos en ese esquema.

## Control de acceso basado en roles (RBAC)

Utilizando el control de acceso basado en roles (RBAC) para administrar los permisos de base de datos en Amazon Redshift, puede simplificar la administración de los permisos de seguridad en Amazon Redshift. Puede proteger el acceso a información confidencial controlando lo que los usuarios pueden hacer tanto en un nivel general como detallado. También puede controlar el acceso de los usuarios a tareas que normalmente solo se permiten a los superusuarios. Asignando permisos diferentes a roles diferentes y asignándolos a usuarios diferentes, puede tener un control más detallado del acceso de los usuarios.

Los usuarios con un rol asignado solo pueden realizar las tareas que estén especificadas por el rol asignado para el que tengan autorización. Por ejemplo, un usuario con el rol asignado que tenga los permisos CREATE TABLE y DROP TABLE solo tiene autorización para realizar esas tareas. Puede controlar el acceso de los usuarios concediendo diferentes niveles de permisos de seguridad a usuarios diferentes para que accedan a los datos que necesitan para su trabajo.

RBAC aplica el principio de permisos mínimos a los usuarios en función de sus requisitos de rol, con independencia de los tipos de objetos involucrados. La concesión y revocación de permisos se realiza en el nivel del rol, sin necesidad de actualizar los permisos en objetos de base de datos individuales.

Con RBAC, puede crear roles con permisos para ejecutar comandos que antes necesitaban permisos de superusuario. Los usuarios pueden ejecutar estos comandos, siempre que hayan recibido autorización con un rol que incluya tales permisos. De manera similar, también puede crear roles para limitar el acceso a determinados comandos y asignar el rol a superusuarios o usuarios que hayan recibido autorización con ese rol.

Para obtener información sobre cómo funciona Amazon Redshift RBAC, vea el siguiente video:Introducción al control de acceso basado en roles (RBAC) en Amazon Redshift.

## Jerarquía de roles

Los roles son conjuntos de permisos que se pueden asignar a un usuario o a otro rol. Puede asignar permisos del sistema o de base de datos a un rol. Un usuario hereda los permisos de un rol asignado.

En RBAC, los usuarios pueden tener roles anidados. Puede conceder roles tanto a usuarios como a roles. Cuando se concede un rol a un usuario, se autoriza a ese usuario con todos los permisos que incluya este rol. Cuando se concede un rol r1 a un usuario, se autoriza a ese usuario con los permisos de r1. El usuario ahora tiene los permisos de r1, así como cualquier otro permiso que ya tuviera.

Cuando se concede un rol (r1) a otro rol (r2), se autoriza a r2 con todos los permisos de r1. Además, cuando se concede r2 a otro rol (r3), los permisos de r3 son la combinación de los permisos de r1 y r2. La jerarquía de roles hace que r2 herede los permisos de r1. Amazon Redshift propaga los permisos con cada autorización de rol. Cuando se concede r1 a r2, y después r2 a r3, se autoriza a r3 con todos los permisos de los tres roles. Por lo tanto, si se concede r3 a un usuario, ese usuario tiene todos los permisos de los tres roles.

Amazon Redshift no permite crear un ciclo de autorización de roles. Un ciclo de autorización de roles ocurre cuando se asigna un rol anidado a un rol anterior en la jerarquía de roles; por ejemplo, asignar r3 a r1. Para obtener más información sobre cómo crear roles y administrar la asignación de roles, consulte Administración de roles en RBAC.

## Asignación de roles

Los superusuarios y los usuarios normales con los permisos CREATE ROLE pueden utilizar la instrucción CREATE ROLE para crear roles. Los superusuarios y los administradores de roles pueden utilizar la instrucción GRANT ROLE para conceder un rol a otros usuarios. Pueden utilizar la instrucción REVOKE ROLE para revocar un rol a otros usuarios, y la instrucción DROP ROLE para eliminar roles. Los administradores de roles incluyen a los propietarios de roles y a los usuarios a los que se les ha concedido el rol con el permiso ADMIN OPTION.

Únicamente los superusuarios o administradores de roles pueden conceder y revocar roles. Se pueden conceder o revocar uno o varios roles a uno o varios roles o usuarios. Utilice la opción WITH ADMIN OPTION de la instrucción GRANT ROLE para proporcionar las opciones de administración de todos los roles concedidos a todos los concesionarios.

Amazon Redshift admite diferentes combinaciones de asignaciones de roles, como conceder varios roles o tener varios concesionarios. WITH ADMIN OPTION solo se aplica a los usuarios, no a los

roles. De manera similar, utilice la opción WITH ADMIN OPTION en la instrucción REVOKE ROLE para eliminar el rol y la autorización administrativa del concesionario. Cuando se utiliza con ADMIN OPTION, solo se revoca la autorización administrativa al rol.

En el siguiente ejemplo, se revoca la autorización administrativa del rol `sample_role2` a `user2`.

```
REVOKE ADMIN OPTION FOR sample_role2 FROM user2;
```

Para obtener más información sobre cómo crear roles y administrar la asignación de roles, consulte Administración de roles en RBAC.

## Roles definidos por el sistema de Amazon Redshift

Amazon Redshift proporciona algunos roles definidos por el sistema que se definen con permisos específicos. Los roles específicos del sistema van precedidos del prefijo `sys:`. Únicamente los usuarios con acceso adecuado pueden modificar los roles definidos por el sistema o crear roles definidos por el sistema personalizados. No se puede utilizar el prefijo `sys:` para un rol definido por el sistema personalizado.

La tabla siguiente resume los roles y sus permisos.

| Nombre del rol | Descripción | | |
|---|---|---|---|
| `sys:monitor` | Este rol tiene permiso para acceder a tablas del catálogo o sistema. | | |
| `sys:operator` | Este rol tiene permisos para acceder a tablas de catálogo o sistema, analizar, limpiar o cancelar consultas. | | |
| `sys:dba` | Este rol tiene permisos para crear esquemas, crear tablas, eliminar | | |

| Nombre del rol | Descripción | | |
|---|---|---|---|
| | esquemas, eliminar tablas y truncar tablas. Tiene permisos para crear o reemplazar procedimientos almacenados, eliminar procedimientos, crear o reemplazar funciones, crear o reemplazar funciones externas, crear vistas y eliminar vistas. Además, este rol hereda todos los permisos del rol sys:operator. | | |
| `sys:superuser` | Este rol tiene todos los permisos del sistema admitidos que se han definido en Permisos del sistema para RBAC. | | |

| Nombre del rol | Descripción | | |
|---|---|---|---|
| `sys:secadmin` | • Este rol tiene permisos para crear usuarios, modificar usuarios, eliminar usuarios, crear roles, eliminar roles y conceder roles.<br>• Este rol tiene permisos para activar o desactiva r el RLS en una relación y permisos para administr ar las políticas de RLS y DDM (CREATE, DROP, ATTACH, DETACH y ALTER). Además, tenga en cuenta que los permisos EXPLAIN RLS, IGNORE RLS y EXPLAIN MASKING se conceden a este rol de forma predeterm inada.<br>• Este rol solo puede tener acceso a tablas de usuario cuando se concede el permiso de forma explícita al rol. | | |

## Roles y usuarios definidos por el sistema para el intercambio de datos

Amazon Redshift crea roles y usuarios para uso interno que corresponden a los recursos compartidos de datos y a sus consumidores. Cada nombre de rol y nombre de usuario internos tienen el prefijo de espacio de nombres reservado `ds:`. Tienen el siguiente formato:

| Nombre | Descripción | | |
|---|---|---|---|
| `ds:`*`sharename`* | Un rol del sistema que se corresponde con un recurso compartido de datos. | | |
| `ds:`*`sharename_consumer`* | Un usuario del sistema que se corresponde con consumidor de recurso compartido de datos. | | |

Se crea un rol de intercambio de datos para cada recurso compartido de datos. Contiene todos los permisos concedidos actualmente al recurso compartido de datos. Se crea un usuario de intercambio de datos para cada consumidor de un recurso compartido de datos. Se le concede permiso para un único rol de intercambio de datos. Un consumidor agregado a varios conjuntos de datos tendrá un usuario de intercambio de datos creado para cada recurso compartido de datos.

Estos usuarios y roles son necesarios para que el intercambio de datos funcione correctamente. No se pueden modificar ni eliminar, y no se puede acceder a ellos ni utilizarlos para ninguna tarea que realicen los clientes. Puede ignorarlos de forma segura. Para obtener más información sobre el uso compartido de datos, consulte Compartir datos entre clústeres en Amazon Redshift.

> **Note**
>
> No puede usar el prefijo `ds:` para crear roles o usuarios definidos por el usuario.

# Permisos del sistema para RBAC

A continuación se presenta una lista de permisos del sistema que puede conceder o revocar a un rol.

| Comando | Debe recibir permiso para ejecutar el comando de una de las siguientes formas. | | |
|---|---|---|---|
| CREAR ROL | • Superusuario.<br>• Usuarios con el permiso CREATE ROLE. | | |
| DROP ROLE | • Superusuario.<br>• Propietario del rol, que es el usuario que creó el rol o un usuario al que se le ha concedido el rol con el permiso WITH ADMIN OPTION. | | |
| CREAR USUARIO | • Superusuario.<br>• Usuarios con el permiso CREATE USER. Estos usuarios no pueden crear superusuarios. | | |
| DROP USER | • Superusuario.<br>• Usuarios con el permiso DROP USER. | | |

| Comando | Debe recibir permiso para ejecutar el comando de una de las siguientes formas. | | |
|---|---|---|---|
| ALTER USER | • Superusuario.<br>• Usuarios con el permiso ALTER USER. Estos usuarios no pueden convertir a usuarios en superusuarios, ni a superusuarios en usuarios.<br>• Usuario actual que desea cambiar su propia contraseña. | | |
| CREATE SCHEMA | • Superusuario.<br>• Usuarios con el permiso CREATE SCHEMA. | | |
| DROP SCHEMA | • Superusuario.<br>• Usuarios con el permiso DROP SCHEMA.<br>• Propietario del esquema. | | |

| Comando | Debe recibir permiso para ejecutar el comando de una de las siguientes formas. | | |
| --- | --- | --- | --- |
| ALTER DEFAULT PRIVILEGES | • Superusuario.<br>• Usuarios con el permiso ALTER DEFAULT PRIVILEGIES.<br>• Usuarios que cambien sus propios permisos de acceso predeterminados.<br>• Usuarios que establezcan permisos para esquemas para los que tienen permisos de acceso. | | |
| ACCESS CATALOG | • Superusuario.<br>• Usuarios con el permiso ACCESS CATALOG. | | |
| ACCESS SYSTEM TABLE | • Superusuario.<br>• Usuarios con el permiso ACCESS SYSTEM TABLE. | | |

| Comando | Debe recibir permiso para ejecutar el comando de una de las siguientes formas. | | |
| --- | --- | --- | --- |
| CREATE TABLE | • Superusuario.<br>• Usuarios con el permiso CREATE TABLE.<br>• Usuarios con el permiso CREATE en esquemas. | | |
| DROP TABLE | • Superusuario.<br>• Usuarios con el permiso DROP TABLE.<br>• Propietario de la tabla con el permiso USAGE en el esquema. | | |
| ALTER TABLE | • Superusuario.<br>• Usuarios con el permiso ALTER TABLE.<br>• Propietario de la tabla con el permiso USAGE en el esquema. | | |

| Comando | Debe recibir permiso para ejecutar el comando de una de las siguientes formas. | | |
| --- | --- | --- | --- |
| CREATE OR REPLACE FUNCTION | • Para CREATE FUNCTION:<br>  • Superusuario.<br>  • Usuarios con el permiso CREATE OR REPLACE FUNCTION.<br>  • Usuarios con el permiso USAGE para el lenguaje.<br>• Para REPLACE FUNCTION:<br>  • Superusuario.<br>  • Usuarios con el permiso CREATE OR REPLACE FUNCTION.<br>  • Propietario de la función. | | |
| CREATE OR REPLACE EXTERNAL FUNCTION | • Superusuario.<br>• Usuarios con el permiso CREATE OR REPLACE EXTERNAL FUNCTION. | | |

| Comando | Debe recibir permiso para ejecutar el comando de una de las siguientes formas. | | |
|---|---|---|---|
| DROP FUNCTION | • Superusuario.<br>• Usuarios con el permiso DROP FUNCTION.<br>• Propietario de la función. | | |
| CREATE OR REPLACE PROCEDURE | • Para CREATE PROCEDURE:<br>  • Superusuario.<br>  • Usuarios con el permiso CREATE OR REPLACE PROCEDURE.<br>  • Usuarios con el permiso USAGE para el lenguaje.<br>• Para REPLACE PROCEDURE:<br>  • Superusuario.<br>  • Usuarios con el permiso CREATE OR REPLACE PROCEDURE.<br>  • Propietario del procedimiento. | | |

| Comando | Debe recibir permiso para ejecutar el comando de una de las siguientes formas. | | |
| --- | --- | --- | --- |
| DROP PROCEDURE | • Superusuario.<br>• Usuarios con el permiso DROP PROCEDURE.<br>• Propietario del procedimiento. | | |
| CREATE OR REPLACE VIEW | • Para CREATE VIEW:<br>  • Superusuario.<br>  • Usuarios con el permiso CREATE OR REPLACE VIEW.<br>  • Usuarios con el permiso CREATE en esquemas.<br>• Para REPLACE VIEW:<br>  • Superusuario.<br>  • Usuarios con el permiso CREATE OR REPLACE VIEW.<br>  • Propietario de la vista. | | |

| Comando | Debe recibir permiso para ejecutar el comando de una de las siguientes formas. | | |
|---|---|---|---|
| DROP VIEW | • Superusuario.<br>• Usuarios con el permiso DROP VIEW.<br>• Propietario de la vista. | | |
| CREATE MODEL | • Superusuario.<br>• Usuarios con el permiso del sistema CREATE MODEL, que deberían poder leer la relación de CREATE MODEL.<br>• Usuarios con el permiso CREATE MODEL. | | |
| DROP MODEL | • Superusuario.<br>• Usuarios con el permiso DROP MODEL.<br>• Propietario del modelo.<br>• Propietario del esquema. | | |

| Comando | Debe recibir permiso para ejecutar el comando de una de las siguientes formas. | | |
| --- | --- | --- | --- |
| CREATE DATASHARE | • Superusuario.<br>• Usuarios con el permiso CREATE DATASHARE.<br>• Propietario de la base de datos. | | |

| Comando | Debe recibir permiso para ejecutar el comando de una de las siguientes formas. | | |
| --- | --- | --- | --- |
| ALTER DATASHARE | • Superusuario.<br>• Usuario con el permiso ALTER DATASHARE.<br>• Usuarios que tienen el permiso ALTER u ALL en el recurso compartido de datos.<br>• Para agregar objetos específicos a un recurso compartido de datos, estos usuarios deben disponer de permisos para los objetos. Los usuarios deben ser propietarios de los objetos o disponer de permisos SELECT, USAGE u ALL para los objetos. | | |

| Comando | Debe recibir permiso para ejecutar el comando de una de las siguientes formas. | | |
| --- | --- | --- | --- |
| DROP DATASHARE | • Superusuario.<br>• Usuarios con el permiso DROP DATASHARE.<br>• Propietario de la base de datos. | | |
| CREATE LIBRARY | • Superusuario.<br>• Usuarios con el permiso CREATE LIBRARY o con el permiso del lenguaje especific ado. | | |
| DROP LIBRARY | • Superusuario.<br>• Usuarios con el permiso DROP LIBRARY.<br>• Propietario de la biblioteca. | | |

| Comando | Debe recibir permiso para ejecutar el comando de una de las siguientes formas. | | |
| --- | --- | --- | --- |
| ANALYZE | • Superusuario.<br>• Usuarios con el permiso ANALYZE.<br>• Propietario de la relación.<br>• Propietario de la base de datos con quien se comparte la tabla. | | |
| CANCEL | • Superusuario que cancele su propia consulta.<br>• Superusuario que cancele la consulta de un usuario.<br>• Usuarios con el permiso CANCEL que cancelen la consulta de un usuario.<br>• Usuario que cancele su propia consulta. | | |

| Comando | Debe recibir permiso para ejecutar el comando de una de las siguientes formas. | | |
|---|---|---|---|
| TRUNCATE TABLE | • Superusuario.<br>• Usuarios con el permiso TRUNCATE TABLE.<br>• Propietario de la tabla. | | |
| VACUUM | • Superusuario.<br>• Usuarios con el permiso VACUUM.<br>• Propietario de la tabla.<br>• Propietario de la base de datos con quien se comparte la tabla. | | |
| IGNORE RLS | • Superusuario.<br>• Usuarios en el rol `sys:secadmin` . | | |
| EXPLAIN RLS | • Superusuario.<br>• Usuarios en el rol `sys:secadmin` . | | |
| EXPLAIN MASKING | • Superusuario.<br>• Usuarios en el rol `sys:secadmin` . | | |

## Permisos de objetos de bases de datos

Además de permisos del sistema, Amazon Redshift incluye permisos de objetos de base de datos que definen las opciones de acceso. Entre esas opciones se incluyen la capacidad de leer datos en tablas y vistas, escribir datos, crear tablas y eliminar tablas. Para obtener más información, consulte GRANT.

Mediante RBAC, se pueden asignar permisos de objetos de base de datos a roles, de manera similar a como se puede hacer con los permisos del sistema. Después, se pueden asignar roles a los usuarios, autorizar a los usuarios con permisos del sistema, y autorizar a los usuarios con permisos de base de datos.

## ALTER DEFAULT PRIVILEGES para RBAC

Utilice la instrucción ALTER DEFAULT PRIVILEGES para definir el conjunto predeterminado de permisos de acceso que se deben aplicar a los objetos que el usuario especificado cree en el futuro. De manera predeterminada, los usuarios pueden cambiar solo sus propios permisos de acceso predeterminados. Con RBAC, se pueden establecer los permisos de acceso predeterminados para los roles. Para obtener más información, consulte el comando ALTER DEFAULT PRIVILEGES.

RBAC permite asignar permisos de objetos de base de datos a los roles, de manera similar a los permisos del sistema. Después, se pueden asignar roles a los usuarios, y autorizar a los usuarios con permisos del sistema o de base de datos.

## Consideraciones sobre el uso de roles en RBAC

Cuando trabaje con roles de RBAC, tenga presente lo siguiente:

- Amazon Redshift no permite ciclos de autorizaciones de roles. No se puede conceder r1 a r2 y luego conceder r2 a r1.
- RBAC funciona tanto para objetos nativos de Amazon Redshift como para tablas de Amazon Redshift Spectrum.
- Como administrador de Amazon Redshift, puede activar RBAC actualizando el clúster al último parche de mantenimiento para comenzar.
- Únicamente los superusuarios y los usuarios con el permiso del sistema CREATE ROLE pueden crear roles.
- Únicamente los superusuarios y administradores de roles pueden modificar o eliminar roles.
- El nombre de un rol no puede ser el mismo que el nombre de un usuario.

- El nombre de un rol no puede contener caracteres no válidos, como ":/\n.".
- El nombre de un rol no puede ser una palabra reservada, como PUBLIC.
- El nombre del rol no puede comenzar con el prefijo reservado para los roles predeterminados, `sys:`.
- No se puede eliminar un rol que tenga el parámetro RESTRICT cuando está concedido a otro rol. La configuración predeterminada es RESTRICT. Amazon Redshift genera un error cuando se intenta eliminar un rol que ha heredado otro rol.
- Los usuarios que no tengan permisos de administración para un rol no pueden conceder ni revocar un rol.
- RBAC no es totalmente compatible con tablas y vistas del sistema. Los permisos de RBAC para las tablas y vistas del sistema no se conservan durante las actualizaciones, las degradaciones o los cambios de tamaño. Recomendamos utilizar Roles definidos por el sistema de Amazon Redshift para administrar los permisos de tablas y vistas del sistema. Para obtener más información sobre las tablas del sistema, vaya a Referencia de las tablas y vistas de sistema.

## Administración de roles en RBAC

Para realizar las siguientes acciones, utilice estos comandos:

- Para crear un rol, utilice el comando CREAR ROL.
- Para cambiar el nombre de un rol o el propietario del rol, utilice el comando ALTER ROLE.
- Para eliminar un rol, utilice el comando DROP ROLE.
- Para conceder un rol a un usuario, utilice el comando GRANT.
- Para revocar un rol a un usuario, utilice el comando REVOKE.
- Para conceder permisos del sistema a un rol, utilice el comando GRANT.
- Para revocar permisos del sistema a un rol, utilice el comando REVOKE.

Para ver una lista de los roles de su clúster o grupo de trabajo, consulte SVV_ROLES.

## Tutorial: creación de roles y realización de consultas con RBAC

Con RBAC, puede crear roles con permisos para ejecutar comandos que antes necesitaban permisos de superusuario. Los usuarios pueden ejecutar estos comandos, siempre que hayan recibido autorización con un rol que incluya tales permisos.

En este tutorial, utilizará el control de acceso basado en roles (RBAC) para administrar los permisos en una base de datos que ha creado. A continuación, se conectará a la base de datos y la consultará desde dos roles diferentes para probar la funcionalidad de RBAC.

Los dos roles que se crean y utilizan para consultar la base de datos son `sales_ro` y `sales_rw`. Cree el rol `sales_ro` y consulte los datos como usuario con el rol `sales_ro`. El usuario `sales_ro` solo puede usar el comando SELECT y no puede usar el comando UPDATE. A continuación, cree el rol `sales_rw` y consulte los datos como usuario con el rol `sales_rw`. El usuario `sales_rw` puede usar el comando SELECT y el comando UPDATE.

Además, puede crear roles para limitar el acceso a determinados comandos y asignar el rol a superusuarios o a usuarios.

Tareas


- Requisitos previos
- Paso 1: cree un usuario administrador
- Paso 2: configure los esquemas
- Paso 3: cree un usuario de solo lectura.
- Paso 4: consulte los datos como usuario de solo lectura
- Paso 5: cree un usuario de solo escritura
- Paso 6: consulte los datos como el usuario con el rol heredado de solo lectura
- Paso 7: conceda permisos de actualización e inserción al rol de lectura/escritura
- Paso 8: consulte los datos como usuario de lectura/escritura
- Paso 9: analice y vacíe las tablas de una base de datos como usuario administrador
- Paso 10: trunque las tablas como usuario de lectura/escritura
- Funciones del sistema para RBAC (opcional)
- Vistas del sistema para RBAC (opcional)
- Uso de seguridad a nivel de fila con RBAC (opcional)


## Requisitos previos

- Cree un clúster de Amazon Redshift o un grupo de trabajo sin servidor que se cargue con la base de datos de ejemplo TICKIT. Para crear un grupo de trabajo sin servidor, consulte Introducción a los almacenamientos de datos de Redshift sin servidor. Para crear un clúster, consulte Crear un

clúster de Amazon Redshift de muestra. Para obtener más información acerca de la base de datos de ejemplo TICKIT, consulte Base de datos de muestra.

- Tenga acceso a un usuario con permisos de superusuario o administrador de roles. Únicamente los superusuarios o administradores de roles pueden conceder o revocar roles. Para obtener más información acerca de los permisos requeridos para RBAC, consulte Permisos del sistema para RBAC.
- Consulte el Consideraciones sobre el uso de roles en RBAC.

## Paso 1: cree un usuario administrador

Para prepararse para este tutorial, cree un rol de administrador de base de datos y asígnelo a un usuario administrador de bases de datos en este paso. Debe crear el administrador de la base de datos como superusuario o administrador de roles.

Ejecute todas las consultas en el editor de consultas v2 de Amazon Redshift.

1. Para crear el rol de administrador db_admin, utilice el siguiente ejemplo.

```
CREATE ROLE db_admin;
```

2. Para crear un usuario de base de datos llamado dbadmin, utilice el siguiente ejemplo.

```
CREATE USER dbadmin PASSWORD 'Test12345';
```

3. Para conceder el rol definido por el sistema denominado sys:dba al rol db_admin, utilice el siguiente ejemplo. Cuando se le concede el rol sys:dba, el usuario dbadmin puede crear esquemas y tablas. Para obtener más información, consulte Roles definidos por el sistema de Amazon Redshift.

## Paso 2: configure los esquemas

En este paso, se conecta a la base de datos como administrador de bases de datos. A continuación, cree dos esquemas y agregue datos en ellos.

1. Conéctese a la base de datos dev como usuario dbadmin utilizando el editor de consultas v2. Para obtener más información sobre la conexión a una base de datos, consulte Trabajo con el editor de consultas v2.
2. Para crear los esquemas de las bases de datos de ventas y marketing, utilice el siguiente ejemplo.

```
CREATE SCHEMA sales;
CREATE SCHEMA marketing;
```

3. Para crear e insertar valores en las tablas del esquema de ventas, utilice el siguiente ejemplo.

```
CREATE TABLE sales.cat(
catid smallint,
catgroup varchar(10),
catname varchar(10),
catdesc varchar(50)
);
INSERT INTO sales.cat(SELECT * FROM category);

CREATE TABLE sales.dates(
dateid smallint,
caldate date,
day char(3),
week smallint,
month char(5),
qtr char(5),
year smallint,
holiday boolean
);
INSERT INTO sales.dates(SELECT * FROM date);

CREATE TABLE sales.events(
eventid integer,
venueid smallint,
catid smallint,
dateid smallint,
eventname varchar(200),
starttime timestamp
);
INSERT INTO sales.events(SELECT * FROM event);

 CREATE TABLE sales.sale(
salesid integer,
listid integer,
sellerid integer,
buyerid integer,
eventid integer,
dateid smallint,
```

```
qtysold smallint,
pricepaid decimal(8,2),
commission decimal(8,2),
saletime timestamp
);
INSERT INTO sales.sale(SELECT * FROM sales);
```

4. Para crear e insertar valores en las tablas del esquema de marketing, utilice el siguiente ejemplo.

```
CREATE TABLE marketing.cat(
catid smallint,
catgroup varchar(10),
catname varchar(10),
catdesc varchar(50)
);
INSERT INTO marketing.cat(SELECT * FROM category);

CREATE TABLE marketing.dates(
dateid smallint,
caldate date,
day char(3),
week smallint,
month char(5),
qtr char(5),
year smallint,
holiday boolean
);
INSERT INTO marketing.dates(SELECT * FROM date);

CREATE TABLE marketing.events(
eventid integer,
venueid smallint,
catid smallint,
dateid smallint,
eventname varchar(200),
starttime timestamp
);
INSERT INTO marketing.events(SELECT * FROM event);

CREATE TABLE marketing.sale(
marketingid integer,
listid integer,
sellerid integer,
buyerid integer,
```

```
eventid integer,
dateid smallint,
qtysold smallint,
pricepaid decimal(8,2),
commission decimal(8,2),
saletime timestamp
);
INSERT INTO marketing.sale(SELECT * FROM marketing);
```

## Paso 3: cree un usuario de solo lectura.

En este paso, creará un rol de solo lectura y un usuario salesanalyst para el rol de solo lectura. El analista de ventas solo necesita acceso de solo lectura a las tablas del esquema de ventas para llevar a cabo la tarea que se le ha asignado de encontrar los eventos que generaron las mayores comisiones.

1. Conéctese a la base de datos como usuario dbadmin.
2. Para crear el rol sales_ro, utilice el siguiente ejemplo.

```
CREATE ROLE sales_ro;
```

3. Para crear el usuario salesanalyst, utilice el siguiente ejemplo.

```
CREATE USER salesanalyst PASSWORD 'Test12345';
```

4. Para conceder el uso del rol sales_ro y seleccionar el acceso a los objetos del esquema de ventas, utilice el siguiente ejemplo.

```
GRANT USAGE ON SCHEMA sales TO ROLE sales_ro;
GRANT SELECT ON ALL TABLES IN SCHEMA sales TO ROLE sales_ro;
```

5. Para conceder al usuario salesanalyst el rol sales_ro, utilice el siguiente ejemplo.

```
GRANT ROLE sales_ro TO salesanalyst;
```

## Paso 4: consulte los datos como usuario de solo lectura

En este paso, el usuario salesanalyst consulta los datos del esquema de ventas. A continuación, el usuario salesanalyst intenta actualizar una tabla y leer las tablas del esquema de marketing.

1. Conéctese a la base de datos como usuario salesanalyst.
2. Para encontrar las diez ventas con las comisiones más altas, utilice el siguiente ejemplo.

```
SET SEARCH_PATH TO sales;
SELECT DISTINCT events.dateid, sale.commission, cat.catname
FROM sale, events, dates, cat
WHERE events.dateid=dates.dateid AND events.dateid=sale.dateid AND events.catid =
 cat.catid
ORDER BY 2 DESC LIMIT 10;

+--------+------------+----------+
| dateid | commission | catname  |
+--------+------------+----------+
|   1880 |     1893.6 | Pop      |
|   1880 |     1893.6 | Opera    |
|   1880 |     1893.6 | Plays    |
|   1880 |     1893.6 | Musicals |
|   1861 |       1500 | Plays    |
|   2003 |       1500 | Pop      |
|   1861 |       1500 | Opera    |
|   2003 |       1500 | Plays    |
|   1861 |       1500 | Musicals |
|   1861 |       1500 | Pop      |
+--------+------------+----------+
```

3. Para seleccionar diez eventos de la tabla de eventos del esquema de ventas, utilice el siguiente ejemplo.

```
SELECT * FROM sales.events LIMIT 10;

+---------+---------+-------+--------+--------------------+---------------------+
| eventid | venueid | catid | dateid |     eventname      |      starttime      |
+---------+---------+-------+--------+--------------------+---------------------+
|    4836 |      73 |     9 |   1871 | Soulfest           | 2008-02-14 19:30:00 |
|    5739 |      41 |     9 |   1871 | Fab Faux           | 2008-02-14 19:30:00 |
|     627 |     229 |     6 |   1872 | High Society       | 2008-02-15 14:00:00 |
|    2563 |     246 |     7 |   1872 | Hamlet             | 2008-02-15 20:00:00 |
|    7703 |      78 |     9 |   1872 | Feist              | 2008-02-15 14:00:00 |
|    7903 |      90 |     9 |   1872 | Little Big Town    | 2008-02-15 19:30:00 |
|    7925 |     101 |     9 |   1872 | Spoon              | 2008-02-15 19:00:00 |
|    8113 |      17 |     9 |   1872 | Santana            | 2008-02-15 15:00:00 |
|     463 |     303 |     8 |   1873 | Tristan und Isolde | 2008-02-16 19:00:00 |
|     613 |     236 |     6 |   1873 | Pal Joey           | 2008-02-16 15:00:00 |
```

```
+---------+---------+-------+--------+--------------------+--------------------+
```

4. Para intentar actualizar el nombre del evento para el eventid 1, ejecute el siguiente ejemplo. Este ejemplo generará un error de permiso denegado porque el usuario salesanalyst solo tiene permisos SELECT en la tabla de eventos del esquema de ventas. Para actualizar la tabla de eventos, debe conceder al rol sales_ro permisos UPDATE. Para obtener más información sobre la concesión de permisos para actualizar una tabla, consulte el parámetro UPDATE para GRANT. Para obtener más información sobre el comando UPDATE, consulte UPDATE.

```
UPDATE sales.events
SET eventname = 'Comment event'
WHERE eventid = 1;

ERROR: permission denied for relation events
```

5. Para intentar seleccionarlos todos en la tabla de eventos del esquema de marketing, use el siguiente ejemplo. Este ejemplo generará un error de permiso denegado porque el usuario salesanalyst solo tiene permisos SELECT en la tabla de eventos del esquema de ventas. Para seleccionar datos de la tabla de eventos del esquema de marketing, debe conceder al rol sales_ro permisos SELECT en la tabla de eventos del esquema de marketing.

```
SELECT * FROM marketing.events;

               ERROR: permission denied for schema marketing
```

## Paso 5: cree un usuario de solo escritura

En este paso, el ingeniero de ventas responsable de crear la canalización de extracción, transformación y carga (ETL) para el procesamiento de datos en el esquema de ventas tendrá acceso de solo lectura, pero más adelante se le dará acceso de lectura y escritura para realizar sus tareas.

1. Conéctese a la base de datos como usuario dbadmin.
2. Para crear el rol sales_rw en el esquema de ventas, utilice el siguiente ejemplo.

```
CREATE ROLE sales_rw;
```

3. Para crear el usuario salesengineer, utilice el siguiente ejemplo.

```
CREATE USER salesengineer PASSWORD 'Test12345';
```

4. Para conceder al rol sales_rw acceso de uso y selección de los objetos del esquema de ventas asignándole el rol sales_ro, utilice el siguiente ejemplo. Para obtener más información sobre cómo los roles heredan los permisos en Amazon Redshift, consulte Jerarquía de roles.

```
GRANT ROLE sales_ro TO ROLE sales_rw;
```

5. Para asignar el rol sales_rw al usuario salesengineer, utilice el siguiente ejemplo.

```
GRANT ROLE sales_rw TO salesengineer;
```

## Paso 6: consulte los datos como el usuario con el rol heredado de solo lectura

En este paso, el usuario salesengineer intenta actualizar la tabla de eventos antes de que se le concedan los permisos de lectura.

1. Conéctese a la base de datos como usuario salesengineer.
2. El usuario salesengineer puede leer correctamente los datos de la tabla de eventos del esquema de ventas. Para seleccionar el evento con eventid 1 de la tabla de eventos del esquema de ventas, utilice el siguiente ejemplo.

```
SELECT * FROM sales.events where eventid=1;

+---------+---------+-------+--------+-----------------+---------------------+
| eventid | venueid | catid | dateid |    eventname    |      starttime      |
+---------+---------+-------+--------+-----------------+---------------------+
|       1 |     305 |     8 |   1851 | Gotterdammerung | 2008-01-25 14:30:00 |
+---------+---------+-------+--------+-----------------+---------------------+
```

3. Para intentar seleccionarlos todos en la tabla de eventos del esquema de marketing, use el siguiente ejemplo. El usuario salesengineer no tiene permisos para las tablas del esquema de marketing, por lo que esta consulta generará un error de permiso denegado. Para seleccionar datos de la tabla de eventos del esquema de marketing, debe conceder al rol sales_rw permisos SELECT en la tabla de eventos del esquema de marketing.

```
SELECT * FROM marketing.events;
```

```
ERROR: permission denied for schema marketing
```

4. Para intentar actualizar el nombre del evento para el eventid 1, ejecute el siguiente ejemplo. Este ejemplo generará un error de permiso denegado porque el usuario salesengineer solo tiene permisos de selección en la tabla de eventos del esquema de ventas. Para actualizar la tabla de eventos, debe conceder al rol sales_rw permisos UPDATE.

```
UPDATE sales.events
SET eventname = 'Comment event'
WHERE eventid = 1;

ERROR: permission denied for relation events
```

## Paso 7: conceda permisos de actualización e inserción al rol de lectura/escritura

En este paso, se conceden permisos de actualización e inserción al rol sales_rw.

1. Conéctese a la base de datos como usuario dbadmin.
2. Para conceder los permisos UPDATE, INSERT y DELETE al rol sales_rw, utilice el siguiente ejemplo.

```
GRANT UPDATE, INSERT, ON ALL TABLES IN SCHEMA sales TO role sales_rw;
```

## Paso 8: consulte los datos como usuario de lectura/escritura

En este paso, salesengineer actualiza correctamente la tabla después de que a su rol se le concedan los permisos de inserción y actualización. A continuación, salesengineer intenta analizar y vaciar la tabla de eventos, pero no lo consigue.

1. Conéctese a la base de datos como usuario salesengineer.
2. Para actualizar el nombre del evento para el eventid 1, ejecute el siguiente ejemplo.

```
UPDATE sales.events
SET eventname = 'Comment event'
WHERE eventid = 1;
```

3. Para ver el cambio realizado en la consulta anterior, utilice el siguiente ejemplo para seleccionar el evento con eventid 1 en la tabla de eventos del esquema de ventas.

```
SELECT * FROM sales.events WHERE eventid=1;

+---------+---------+-------+--------+--------------+--------------------+
| eventid | venueid | catid | dateid |   eventname  |      starttime     |
+---------+---------+-------+--------+--------------+--------------------+
|       1 |     305 |     8 |   1851 | Comment event | 2008-01-25 14:30:00 |
+---------+---------+-------+--------+--------------+--------------------+
```

4. Para analizar la tabla de eventos actualizada en el esquema de ventas, utilice el siguiente ejemplo. Este ejemplo generará un error de permiso denegado porque el usuario salesengineer no tiene los permisos necesarios y no es el propietario de la tabla de eventos del esquema de ventas. Para analizar la tabla de eventos, debe conceder al rol sales_rw permisos ANALYZE mediante el comando GRANT. Para obtener más información acerca del comando ANALYZE, consulte ANALYZE.

```
ANALYZE sales.events;

               ERROR: skipping "events" --- only table or database owner can analyze
```

5. Para vaciar la tabla de eventos actualizada, utilice el siguiente ejemplo. Este ejemplo generará un error de permiso denegado porque el usuario salesengineer no tiene los permisos necesarios y no es el propietario de la tabla de eventos del esquema de ventas. Para vaciar la tabla de eventos, debe conceder permisos VACUUM al rol sales_rw mediante el comando GRANT. Para obtener más información acerca del comando VACUUM, consulte VACUUM.

```
VACUUM sales.events;

ERROR: skipping "events" --- only table or database owner can vacuum it
```

## Paso 9: analice y vacíe las tablas de una base de datos como usuario administrador

En este paso, el usuario dbadmin analiza y vacía todas las tablas. El usuario tiene permisos de administrador en esta base de datos, por lo que puede ejecutar estos comandos.

1. Conéctese a la base de datos como usuario dbadmin.
2. Para analizar la tabla de eventos en el esquema de ventas, utilice el siguiente ejemplo.

```
ANALYZE sales.events;
```

3. Para vaciar la tabla de eventos en el esquema de ventas, utilice el siguiente ejemplo.

```
VACUUM sales.events;
```

4. Para analizar la tabla de eventos en el esquema de marketing, utilice el siguiente ejemplo.

```
ANALYZE marketing.events;
```

5. Para vaciar la tabla de eventos en el esquema de marketing, utilice el siguiente ejemplo.

```
VACUUM marketing.events;
```

## Paso 10: trunque las tablas como usuario de lectura/escritura

En este paso, el usuario salesengineer intenta truncar la tabla de eventos en el esquema de ventas, pero solo lo consigue cuando el usuario dbadmin le concede los permisos de truncamiento.

1. Conéctese a la base de datos como usuario salesengineer.
2. Para intentar eliminar todas las filas de la tabla de eventos del esquema de ventas, utilice el siguiente ejemplo. Este ejemplo generará un error porque el usuario salesengineer no tiene los permisos necesarios y no es el propietario de la tabla de eventos del esquema de ventas. Para truncar la tabla de eventos, debe conceder al rol sales_rw permisos TRUNCATE mediante el comando GRANT. Para obtener más información sobre el comando TRUNCATE, consulte TRUNCATE.

```
TRUNCATE sales.events;

ERROR: must be owner of relation events
```

3. Conéctese a la base de datos como usuario dbadmin.
4. Para conceder privilegios de truncado de tablas al rol sales_rw, utilice el siguiente ejemplo.

```
GRANT TRUNCATE TABLE TO role sales_rw;
```

5. Conéctese a la base de datos como usuario salesengineer utilizando el editor de consultas v2.
6. Para leer los primeros diez eventos de la tabla de eventos del esquema de ventas, utilice el siguiente ejemplo.

```
SELECT * FROM sales.events ORDER BY eventid LIMIT 10;

+---------+---------+-------+--------+----------------------------
+---------------------+
| eventid | venueid | catid | dateid |          eventname          |      starttime
    |
+---------+---------+-------+--------+----------------------------
+---------------------+
|       1 |     305 |     8 |   1851 | Comment event               | 2008-01-25
 14:30:00 |
|       2 |     306 |     8 |   2114 | Boris Godunov               | 2008-10-15
 20:00:00 |
|       3 |     302 |     8 |   1935 | Salome                      | 2008-04-19
 14:30:00 |
|       4 |     309 |     8 |   2090 | La Cenerentola (Cinderella) | 2008-09-21
 14:30:00 |
|       5 |     302 |     8 |   1982 | Il Trovatore                | 2008-06-05
 19:00:00 |
|       6 |     308 |     8 |   2109 | L Elisir d Amore            | 2008-10-10
 19:30:00 |
|       7 |     309 |     8 |   1891 | Doctor Atomic               | 2008-03-06
 14:00:00 |
|       8 |     302 |     8 |   1832 | The Magic Flute             | 2008-01-06
 20:00:00 |
|       9 |     308 |     8 |   2087 | The Fly                     | 2008-09-18
 19:30:00 |
|      10 |     305 |     8 |   2079 | Rigoletto                   | 2008-09-10
 15:00:00 |
+---------+---------+-------+--------+----------------------------
+---------------------+
```

7. Para truncar la tabla de eventos en el esquema de ventas, utilice el siguiente ejemplo.

```
TRUNCATE sales.events;
```

8. Para leer los datos de la tabla de eventos del esquema de ventas, utilice el siguiente ejemplo.

```
SELECT * FROM sales.events ORDER BY eventid LIMIT 10;

+---------+---------+-------+--------+----------------------------
+---------------------+
```

```
| eventid | venueid | catid | dateid |          eventname          |      starttime
    |
+---------+---------+-------+--------+-----------------------------
+---------------------+
```

Creación de roles de solo lectura y de lectura/escritura para el esquema de marketing (opcional)

En este paso, se crean roles de solo lectura y de lectura/escritura para el esquema de marketing.

1. Conéctese a la base de datos como usuario dbadmin.
2. Para crear roles de solo lectura y lectura/escritura para el esquema de marketing, utilice el siguiente ejemplo.

```
CREATE ROLE marketing_ro;

CREATE ROLE marketing_rw;

GRANT USAGE ON SCHEMA marketing TO ROLE marketing_ro, ROLE marketing_rw;

GRANT SELECT ON ALL TABLES IN SCHEMA marketing TO ROLE marketing_ro;

GRANT ROLE marketing_ro TO ROLE marketing_rw;

GRANT INSERT, UPDATE, DELETE ON ALL TABLES IN SCHEMA marketing TO ROLE marketing_rw;

CREATE USER marketinganalyst PASSWORD 'Test12345';

CREATE USER marketingengineer PASSWORD 'Test12345';

GRANT ROLE marketing_ro TO marketinganalyst;

GRANT ROLE marketing_rw TO marketingengineer;
```

## Funciones del sistema para RBAC (opcional)

Amazon Redshift tiene dos funciones para proporcionar información del sistema sobre la pertenencia de los usuarios y la pertenencia de los roles a grupos o roles adicionales: role_is_member_of y user_is_member_of. Estas funciones están disponibles para los superusuarios y los usuarios normales. Los superusuarios pueden comprobar todas las pertenencias de los roles. Los usuarios normales solo pueden comprobar la pertenencia de los roles a los que se les han concedido acceso.

Para usar la función role_is_member_of

1. Conéctese a la base de datos como usuario salesengineer.
2. Para comprobar si el rol sales_rw es miembro del rol sales_ro, utilice el siguiente ejemplo.

```
SELECT role_is_member_of('sales_rw', 'sales_ro');

+-------------------+
| role_is_member_of |
+-------------------+
| true              |
+-------------------+
```

3. Para comprobar si el rol sales_ro es miembro del rol sales_rw, utilice el siguiente ejemplo.

```
SELECT role_is_member_of('sales_ro', 'sales_rw');

+-------------------+
| role_is_member_of |
+-------------------+
| false             |
+-------------------+
```

Para usar la función user_is_member_of

1. Conéctese a la base de datos como usuario salesengineer.
2. En el siguiente ejemplo, se intenta comprobar la pertenencia del usuario salesanalyst. Esta consulta produce un error porque salesengineer no tiene acceso a salesanalyst. Para ejecutar este comando correctamente, conéctese a la base de datos como usuario salesanalyst y utilice el ejemplo.

```
SELECT user_is_member_of('salesanalyst', 'sales_ro');

ERROR
```

3. Conéctese a la base de datos como superusuario.
4. Para comprobar la pertenencia del usuario salesanalyst cuando está conectado como superusuario, utilice el siguiente ejemplo.

```
SELECT user_is_member_of('salesanalyst', 'sales_ro');

+-------------------+
| user_is_member_of |
+-------------------+
| true              |
+-------------------+
```

5. Conéctese a la base de datos como usuario dbadmin.
6. Para comprobar la pertenencia del usuario salesengineer, utilice el siguiente ejemplo.

```
SELECT user_is_member_of('salesengineer', 'sales_ro');

+-------------------+
| user_is_member_of |
+-------------------+
| true              |
+-------------------+

SELECT user_is_member_of('salesengineer', 'marketing_ro');

+-------------------+
| user_is_member_of |
+-------------------+
| false             |
+-------------------+

SELECT user_is_member_of('marketinganalyst', 'sales_ro');

+-------------------+
| user_is_member_of |
+-------------------+
| false             |
+-------------------+
```

## Vistas del sistema para RBAC (opcional)

Para ver los roles, la asignación de roles a los usuarios, la jerarquía de roles y los privilegios de los objetos de la base de datos a través de los roles, utilice las vistas del sistema de Amazon Redshift. Estas vistas están disponibles para los superusuarios y los usuarios normales. Los superusuarios

pueden comprobar todos los detalles de los roles. Los usuarios normales solo pueden comprobar los detalles de los roles a los que se les ha concedido acceso.

1. Para ver una lista de los usuarios a los que se han concedido explícitamente roles en el clúster, utilice el siguiente ejemplo.

```
SELECT * FROM svv_user_grants;
```

2. Para ver una lista de los roles a los que se han concedido explícitamente roles en el clúster, utilice el siguiente ejemplo.

```
SELECT * FROM svv_role_grants;
```

Para ver la lista completa de vistas del sistema, consulte Vistas de metadatos SVV.

## Uso de seguridad a nivel de fila con RBAC (opcional)

Para tener un control de acceso detallado de los datos sensibles, utilice la seguridad a nivel de fila (RLS). Para obtener más información sobre RLS, consulte Seguridad de nivel básico.

En esta sección, creará una política RLS que otorga al usuario `salesengineer` permisos para ver solo las filas de la tabla `cat` que tienen el valor `catdesc` de Major League Baseball. A continuación, consulte la base de datos como usuario `salesengineer`.

1. Conéctese a la base de datos como usuario `salesengineer`.
2. Para ver las cinco primeras entradas de la tabla `cat`, utilice el siguiente ejemplo.

```
SELECT *
FROM sales.cat
ORDER BY catid ASC
LIMIT 5;

+-------+----------+---------+---------------------------------+
| catid | catgroup | catname |             catdesc             |
+-------+----------+---------+---------------------------------+
|     1 | Sports   | MLB     | Major League Baseball           |
|     2 | Sports   | NHL     | National Hockey League          |
|     3 | Sports   | NFL     | National Football League        |
|     4 | Sports   | NBA     | National Basketball Association |
|     5 | Sports   | MLS     | Major League Soccer             |
```

```
+-------+----------+---------+--------------------------------+
```

3. Conéctese a la base de datos como usuario `dbadmin`.
4. Para crear una política de RLS para la columna `catdesc` de la tabla `cat`, utilice el siguiente ejemplo.

```
CREATE RLS POLICY policy_mlb_engineer
WITH (catdesc VARCHAR(50))
USING (catdesc = 'Major League Baseball');
```

5. Para asociar la política de RLS al rol de `sales_rw`, utilice el siguiente ejemplo.

```
ATTACH RLS POLICY policy_mlb_engineer ON sales.cat TO ROLE sales_rw;
```

6. Para modificar la tabla y activar RLS, utilice el siguiente ejemplo.

```
ALTER TABLE sales.cat ROW LEVEL SECURITY ON;
```

7. Conéctese a la base de datos como usuario `salesengineer`.
8. Para intentar ver las cinco primeras entradas de la tabla `cat`, utilice el siguiente ejemplo. Tenga en cuenta que las entradas solo aparecen cuando la columna `catdesc` es `Major League Baseball`.

```
SELECT *
FROM sales.cat
ORDER BY catid ASC
LIMIT 5;

+-------+----------+---------+----------------------+
| catid | catgroup | catname |        catdesc       |
+-------+----------+---------+----------------------+
|     1 | Sports   | MLB     | Major League Baseball |
+-------+----------+---------+----------------------+
```

9. Conéctese a la base de datos como usuario `salesanalyst`.
10. Para intentar ver las cinco primeras entradas de la tabla `cat`, utilice el siguiente ejemplo. Tenga en cuenta que no aparece ninguna entrada porque se aplica la política predeterminada de denegación total.

```
SELECT *
```

```
FROM sales.cat
ORDER BY catid ASC
LIMIT 5;

+-------+----------+---------+----------------------+
| catid | catgroup | catname |        catdesc       |
+-------+----------+---------+----------------------+
```

11. Conéctese a la base de datos como usuario `dbadmin`.
12. Para conceder el permiso IGNORE de RLS al rol `sales_ro`, utilice el siguiente ejemplo. Aquí se conceden al usuario `salesanalyst` los permisos para ignorar las políticas de RLS, ya que es miembro del rol `sales_ro`.

```
GRANT IGNORE RLS TO ROLE sales_ro;
```

13. Conéctese a la base de datos como usuario `salesanalyst`.
14. Para ver las cinco primeras entradas de la tabla `cat`, utilice el siguiente ejemplo.

```
SELECT *
FROM sales.cat
ORDER BY catid ASC
LIMIT 5;

+-------+----------+---------+---------------------------------+
| catid | catgroup | catname |             catdesc             |
+-------+----------+---------+---------------------------------+
|     1 | Sports   | MLB     | Major League Baseball           |
|     2 | Sports   | NHL     | National Hockey League          |
|     3 | Sports   | NFL     | National Football League        |
|     4 | Sports   | NBA     | National Basketball Association |
|     5 | Sports   | MLS     | Major League Soccer             |
+-------+----------+---------+---------------------------------+
```

15. Conéctese a la base de datos como usuario `dbadmin`.
16. Para revocar el permiso IGNORE de RLS del rol `sales_ro`, utilice el siguiente ejemplo.

```
REVOKE IGNORE RLS FROM ROLE sales_ro;
```

17. Conéctese a la base de datos como usuario `salesanalyst`.

18. Para intentar ver las cinco primeras entradas de la tabla `cat`, utilice el siguiente ejemplo. Tenga en cuenta que no aparece ninguna entrada porque se aplica la política predeterminada de denegación total.

```
SELECT *
FROM sales.cat
ORDER BY catid ASC
LIMIT 5;

+-------+----------+---------+----------------------+
| catid | catgroup | catname |        catdesc       |
+-------+----------+---------+----------------------+
```

19. Conéctese a la base de datos como usuario `dbadmin`.

20. Para desconectar la política de RLS de la tabla `cat`, utilice el siguiente ejemplo.

```
DETACH RLS POLICY policy_mlb_engineer ON cat FROM ROLE sales_rw;
```

21. Conéctese a la base de datos como usuario `salesanalyst`.

22. Para intentar ver las cinco primeras entradas de la tabla `cat`, utilice el siguiente ejemplo. Tenga en cuenta que no aparece ninguna entrada porque se aplica la política predeterminada de denegación total.

```
SELECT *
FROM sales.cat
ORDER BY catid ASC
LIMIT 5;

+-------+----------+---------+---------------------------------+
| catid | catgroup | catname |             catdesc             |
+-------+----------+---------+---------------------------------+
|     1 | Sports   | MLB     | Major League Baseball           |
|     2 | Sports   | NHL     | National Hockey League          |
|     3 | Sports   | NFL     | National Football League        |
|     4 | Sports   | NBA     | National Basketball Association |
|     5 | Sports   | MLS     | Major League Soccer             |
+-------+----------+---------+---------------------------------+
```

23. Conéctese a la base de datos como usuario `dbadmin`.

24. Para eliminar la política de RLS, utilice el siguiente ejemplo.

```
DROP RLS POLICY policy_mlb_engineer;
```

25Para eliminar RLS, utilice el siguiente ejemplo.

```
ALTER TABLE cat ROW LEVEL SECURITY OFF;
```

### Temas relacionados de

Para obtener más información sobre RBAC, consulte la siguiente documentación:

- Jerarquía de roles
- Asignación de roles
- Permisos de objetos de bases de datos
- ALTER DEFAULT PRIVILEGES para RBAC

## Seguridad de nivel básico

Con la seguridad de la fila (RLS) en Amazon Redshift, puede tener un control de acceso pormenorizado de los datos confidenciales. Puede decidir qué usuarios o roles pueden acceder a registros de datos específicos dentro de los esquemas o tablas, según las políticas de seguridad que se definen en los objetos de la base de datos. Además de la seguridad de la columna, donde puede conceder permisos a los usuarios a un subconjunto de columnas, utilice políticas de RLS para restringir aún más el acceso a determinadas filas de las columnas visibles. Para obtener más información acerca de la seguridad de la columna, consulte Notas de uso para el control de acceso de nivel de columna.

Al aplicar políticas de RLS en las tablas, puede restringir los conjuntos de los resultados devueltos cuando los usuarios ejecutan consultas.

Al crear políticas de RLS, puede especificar expresiones que determinen si Amazon Redshift devuelve alguna fila existente en la tabla de una consulta. Cuando se crean políticas de RLS para limitar el acceso, no es necesario agregar ni externalizar condiciones adicionales en las consultas.

Le recomendamos que, al crear políticas de RLS, cree políticas simples y evite instrucciones complejas. Al definir políticas de RLS, no utilice en la definición uniones de tablas excesivas que se basen en políticas.

Cuando una política hace referencia a una tabla de búsqueda, Amazon Redshift analiza la tabla adicional, además de la tabla en la que se encuentra la política. Existirán diferencias de rendimiento en la misma consulta para un usuario con una política de RLS adjunta y un usuario sin ninguna política adjunta.

## Uso de políticas de RLS en instrucciones SQL

Al utilizar políticas de RLS en instrucciones SQL, Amazon Redshift aplica las siguientes reglas:

- Amazon Redshift aplica políticas de RLS a las instrucciones SELECT, UPDATE y DELETE de forma predeterminada.
- Con SELECT y UNLOAD, Amazon Redshift filtra las filas según la política definida.
- Con UPDATE, Amazon Redshift actualiza solo las filas que puede ver. Si una política restringe un subconjunto de las filas de una tabla, no podrá actualizarlas.
- Con DELETE, puede eliminar solo las filas que puede ver. Si una política restringe un subconjunto de filas de una tabla, no podrá eliminarlas. Con TRUNCATE, aún puede truncar la tabla.
- Con CREATE TABLE LIKE, las tablas creadas con las opciones LIKE no heredarán la configuración de permisos de la tabla de origen. Del mismo modo, la tabla de destino no heredará las políticas de RLS de la tabla de origen.

## Combinación de varias políticas por usuario

RLS en Amazon Redshift permite asociar varias políticas por usuario y objeto. Si se han definido varias políticas para un usuario, Amazon Redshift aplica todas las políticas con la sintaxis AND u OR en función de la configuración de RLS CONJUNCTION TYPE para la tabla. Para obtener más información acerca del tipo de conjunción, consulte ALTER TABLE.

En una tabla, se pueden asociar varias políticas con el usuario. Puede suceder que varias políticas estén directamente asociadas al usuario o que este pertenezca a varios roles, y los roles tengan diferentes políticas asociadas a ellos.

Cuando varias políticas deben restringir el acceso a las filas en una relación determinada, puede establecer el RLS CONJUNCTION TYPE de la relación en AND. Considere el siguiente ejemplo. Alice solo puede ver los eventos deportivos que tengan el catname NBA, tal como se especifica en la política.

```
-- Create an analyst role and grant it to a user named Alice.
```

```
CREATE ROLE analyst;
CREATE USER alice WITH PASSWORD 'Name_is_alice_1';
GRANT ROLE analyst TO alice;

-- Create an RLS policy that only lets the user see sports.
CREATE RLS POLICY policy_sports
WITH (catgroup VARCHAR(10))
USING (catgroup = 'Sports');

-- Create an RLS policy that only lets the user see NBA.
CREATE RLS POLICY policy_nba
WITH (catname VARCHAR(10))
USING (catname = 'NBA');

-- Attach both to the analyst role.
ATTACH RLS POLICY policy_sports ON category TO ROLE analyst;
ATTACH RLS POLICY policy_nba ON category TO ROLE analyst;

-- Activate RLS on the category table with AND CONJUNCTION TYPE.
ALTER TABLE category ROW LEVEL SECURITY ON CONJUNCTION TYPE AND;

-- Change session to Alice.
SET SESSION AUTHORIZATION alice;

-- Select all from the category table.
SELECT catgroup, catname
FROM category;

 catgroup | catname
---------+---------
 Sports   | NBA
(1 row)
```

Si las distintas políticas deben permitir a los usuarios ver más filas en una relación determinada, el usuario puede establecer el RLS CONJUNCTION TYPE de la relación en OR. Considere el siguiente ejemplo. Alice solo puede ver conciertos y deportes, según lo especificado en la política.

```
-- Create an analyst role and grant it to a user named Alice.
CREATE ROLE analyst;
CREATE USER alice WITH PASSWORD 'Name_is_alice_1';
GRANT ROLE analyst TO alice;

-- Create an RLS policy that only lets the user see concerts.
```

```
CREATE RLS POLICY policy_concerts
WITH (catgroup VARCHAR(10))
USING (catgroup = 'Concerts');

-- Create an RLS policy that only lets the user see sports.
CREATE RLS POLICY policy_sports
WITH (catgroup VARCHAR(10))
USING (catgroup = 'Sports');

-- Attach both to the analyst role.
ATTACH RLS POLICY policy_concerts ON category TO ROLE analyst;
ATTACH RLS POLICY policy_sports ON category TO ROLE analyst;

-- Activate RLS on the category table with OR CONJUNCTION TYPE.
ALTER TABLE category ROW LEVEL SECURITY ON CONJUNCTION TYPE OR;

-- Change session to Alice.
SET SESSION AUTHORIZATION alice;

-- Select all from the category table.
SELECT catgroup, count(*)
FROM category
GROUP BY catgroup ORDER BY catgroup;

 catgroup | count
---------+-------
 Concerts |  3
 Sports   |  5
(2 rows)
```

## Propiedad y administración de la política de RLS

Como superusuario, administrador de seguridad o usuario que tiene el rol sys:secadmin, puede crear, modificar, asociar y desasociar políticas de RLS. Las políticas de RLS se pueden asociar a tablas, vistas, vistas de enlace tardío (LBV) y vistas materializadas (MV). En el objeto, puede activar o desactivar la seguridad de la fila sin modificar la definición de esquema de las tablas.

Para comenzar con la seguridad de la fila, a continuación se indican las instrucciones SQL que puede utilizar:

- Utilice la instrucción ALTER TABLE para activar o desactivar RLS en una tabla, vista o vista de enlace tardío. Para obtener más información, consulte ALTER TABLE.

- La instrucción ALTER MATERIALIZED VIEW se utiliza para activar o desactivar RLS en una vista materializada (MV). Para obtener más información, consulte ALTER MATERIALIZED VIEW.
- La instrucción CREATE RLS POLICY se utiliza para crear una política de seguridad para una o más tablas y especificar uno o más usuarios o roles en la política.

  Para obtener más información, consulte CREATE RLS POLICY.
- Utilice la instrucción ALTER RLS POLICY para modificar la política, por ejemplo, cambiar la definición de la política. Puede utilizar la misma política para varias tablas o vistas.

  Para obtener más información, consulte ALTER RLS POLICY.
- La instrucción ATTACH RLS POLICY se utiliza para adjuntar una política a una o más relaciones, a uno o más usuarios, o a roles.

  Para obtener más información, consulte ATTACH RLS POLICY.
- La instrucción DETACH RLS POLICY se utiliza para desconectar una política de una o más relaciones, de uno o más usuarios o de roles.

  Para obtener más información, consulte DETACH RLS POLICY.
- La instrucción DROP RLS POLICY se utiliza para eliminar una política.

  Para obtener más información, consulte DROP RLS POLICY.
- Las instrucciones GRANT y REVOKE se utilizan para conceder y revocar de manera explícita permisos SELECT a las políticas de RLS que hacen referencia a tablas de búsqueda. Para obtener más información, consulte GRANT y REVOKE.

Para supervisar las políticas creadas, sys:secadmin puede ver la SVV_RLS_POLICY y la SVV_RLS_ATTACHED_POLICY.

Para enumerar las relaciones protegidas por RLS, sys:secadmin puede ver SVV_RLS_RELATION.

Para rastrear la aplicación de políticas de RLS en consultas que hacen referencia a relaciones protegidas por RLS, un superusuario, sys:operator, o cualquier usuario con el permiso del sistema ACCESS SYSTEM TABLE puede ver la SVV_RLS_APPLIED_POLICY. Tenga en cuenta que no se concede estos permisos a sys:secadmin de forma predeterminada.

Para permitir a los usuarios el acceso completo a una relación protegida por RLS, puede conceder el permiso IGNORE RLS. A los superusuarios o sys:secadmin se les concede automáticamente IGNORE RLS. Para obtener más información, consulte GRANT.

Para explicar los filtros de política de RLS de una consulta en el plan EXPLAIN a fin de solucionar problemas de consultas relacionadas con RLS, puede conceder el permiso EXPLAIN RLS a cualquier usuario. Para obtener más información, consulte GRANT y EXPLAIN.

## Objetos y principios dependientes de políticas

Para proporciona seguridad a las aplicaciones y evitar que los objetos de política queden obsoletos o no sean válidos, Amazon Redshift no permite eliminar ni alterar los objetos a los que hagan referencia las políticas de RLS.

A continuación, se enumeran las dependencias de objetos de esquema de los que Amazon Redshift hace un seguimiento para las políticas de RLS.

- Al realizar el seguimiento de la dependencia de objetos de esquema para la tabla de destino, Amazon Redshift sigue estas reglas:
  - Amazon Redshift desconecta la política de una relación, usuario, rol o público cuando se elimina una tabla de destino.
  - Si se cambia el nombre de una tabla de destino, no se produce ningún impacto en las políticas adjuntas.
  - Solo se pueden eliminar las columnas de la tabla de destino a la que se hace referencia dentro de la definición de política si se elimina o desconecta primero la política. Esto también se aplica cuando se especifica la opción CASCADE. Puede eliminar otras columnas de la tabla de destino.
  - No se puede cambiar el nombre de las columnas de la tabla de destino a las que se hace referencia. Para cambiar el nombre de las columnas a las que se hace referencia, desconecte primero la política. Esto también se aplica cuando se especifica la opción CASCADE.
  - No se puede cambiar el tipo de la columna a la que se hace referencia, aunque especifique la opción CASCADE.
- Al realizar el seguimiento de la dependencia de objetos de esquema para la tabla de búsqueda, Amazon Redshift sigue estas reglas:
  - No se puede eliminar una tabla de búsqueda. Para eliminar una tabla de búsqueda, primero se debe eliminar la política en la que se hace referencia a la tabla de búsqueda.
  - No se puede cambiar el nombre de una tabla de búsqueda. Para cambiar el nombre de una tabla de búsqueda, primero se debe eliminar la política en la que se hace referencia a la tabla de búsqueda. Esto también se aplica cuando se especifica la opción CASCADE.

  - No se pueden eliminar las columnas de la tabla de búsqueda utilizadas en la definición de la política. Para eliminar las columnas de la tabla de búsqueda utilizadas en la definición de la política, primero se debe eliminar la política en la que se hace referencia a la tabla de búsqueda. Esto también se aplica cuando se especifica la opción CASCADE en la instrucción ALTER TABLE DROP COLUMN. Puede eliminar otras columnas de la tabla de búsqueda.
  - No se puede cambiar el nombre de las columnas de la tabla de búsqueda a las que se hace referencia. Para cambiar el nombre de las columnas referidas, primero se debe eliminar la política en la que se hace referencia a la tabla de búsqueda. Esto también se aplica cuando se especifica la opción CASCADE.
  - No puede cambiar el tipo de la columna a la que se hace referencia.
- Cuando se elimina un usuario o rol, Amazon Redshift desconecta todas las políticas adjuntas al usuario o rol de manera automática.
- Cuando se utiliza la opción CASCADE en la instrucción DROP SCHEMA, Amazon Redshift también elimina las relaciones en el esquema. Además, elimina las relaciones de otros esquemas que dependan de las relaciones del esquema eliminado. En caso de una relación de tabla de búsqueda en una política, Amazon Redshift no pasa el DDL DROP SCHEMA. Amazon Redshift desconecta todas las políticas que se adjuntan a las relaciones eliminadas por la instrucción DROP SCHEMA.
- Solo se puede eliminar una función de búsqueda (una función a la que se hace referencia dentro de una definición de política) cuando también se elimina la política. Esto también se aplica cuando se especifica la opción CASCADE.
- Cuando se adjunta una política a una tabla, Amazon Redshift comprueba si esta tabla es una tabla de búsqueda de una política diferente. En caso afirmativo, Amazon Redshift no permite adjuntar una política a esta tabla.
- Al crear una política de RLS, Amazon Redshift comprueba si esta tabla es una tabla de destino de otra política de RLS. En caso afirmativo, Amazon Redshift no permite crear una política en esta tabla.

## Ejemplo

En el siguiente ejemplo, se ilustra cómo se hace un seguimiento de la dependencia del esquema.

```
-- The CREATE and ATTACH policy statements for `policy_events` references some
-- target and lookup tables.
-- Target tables are tickit_event_redshift and target_schema.target_event_table.
-- Lookup table is tickit_sales_redshift.
```

```
-- Policy `policy_events` has following dependencies:
--   table tickit_sales_redshift column eventid, qtysold
--   table tickit_event_redshift column eventid
--   table target_event_table column eventid
--   schema public and target_schema
CREATE RLS POLICY policy_events
WITH (eventid INTEGER)
USING (
    eventid IN (SELECT eventid FROM tickit_sales_redshift WHERE qtysold <3)
);

ATTACH RLS POLICY policy_events ON tickit_event_redshift TO ROLE analyst;

ATTACH RLS POLICY policy_events ON target_schema.target_event_table TO ROLE consumer;
```

# Consideraciones y limitaciones sobre el uso de las políticas de RLS

## Consideraciones

A continuación, se indican los aspectos que se deben tener en cuenta para trabajar con políticas de RLS:

- Amazon Redshift aplica políticas de RLS a las instrucciones SELECT, UPDATE o DELETE.
- Amazon Redshift no aplica políticas de RLS a las instrucciones INSERT, COPY, ALTER TABLE APPEND.
- Las políticas de RLS se pueden asociar a tablas, vistas, vistas de enlace tardío (LBV) y vistas materializadas (MV).
- La seguridad de la fila funciona con la seguridad de la columna para proteger los datos.
- Cuando se activa RLS para la relación de origen, Amazon Redshift admite la instrucción ALTER TABLE APPEND para los superusuarios, los usuarios a los que se les otorgó de forma explícita el permiso del sistema IGNORE RLS o el rol sys:secadmin. En este caso, con la instrucción ALTER TABLE APPEND, puede mover datos de una tabla de origen existente para anexar filas en una tabla de destino. Amazon Redshift mueve todas las tuplas de la relación de origen a la relación de destino. El estado de RLS de la relación de destino no afecta a la instrucción ALTER TABLE APPEND.
- Para facilitar la migración desde otros sistemas de almacenamiento de datos, puede configurar y recuperar variables de contexto de sesión personalizadas de una conexión especificando el nombre y el valor de la variable.

En el siguiente ejemplo, se establecen variables de contexto de sesión para una política de seguridad de la fila (RLS).

```
-- Set a customized context variable.
SELECT set_config('app.category', 'Concerts', FALSE);

-- Create a RLS policy using current_setting() to get the value of a customized
 context variable.
CREATE RLS POLICY policy_categories
WITH (catgroup VARCHAR(10))
USING (catgroup = current_setting('app.category', FALSE));

-- Set correct roles and attach the policy on the target table to one or more roles.
ATTACH RLS POLICY policy_categories ON tickit_category_redshift TO ROLE analyst, ROLE
 dbadmin;
```

Para obtener más detalles sobre cómo configurar y recuperar variables de contexto de sesión personalizadas, vaya a SET, SET_CONFIG, SHOW, CURRENT_SETTING y RESET. Para obtener más información sobre la modificación de la configuración del servidor en general, vaya a Modificación de la configuración del servidor.

> **Important**
>
> Cuando se utilizan variables de contexto de sesión en las políticas de RLS, la política de seguridad depende del usuario o rol que la invoca. Tenga cuidado de evitar vulnerabilidades de seguridad al utilizar variables de contexto de sesión en las políticas de RLS.

- El cambio de usuario de sesión mediante SET SESSION AUTHORIZATION entre DECLARE y FETCH o entre instrucciones FETCH posteriores no actualizará el plan ya preparado basado en las políticas de usuario en el momento de DECLARE. Evite cambiar de usuario de sesión cuando se utilicen cursores con tablas protegidas por RLS.
- Cuando los objetos base que se encuentran en un objeto de vista están protegidos por RLS, las políticas asociadas al usuario que ejecuta la consulta se aplican a los objetos base respectivos. Esto es diferente de las comprobaciones de permisos en el nivel de objeto, donde los permisos del propietario de la vista se comprueban con los objetos base de la vista. Puede ver las relaciones protegidas por RLS de una consulta en su resultado del plan EXPLAIN.

- Cuando se hace referencia a una función definida por el usuario (UDF) en una política de RLS de una relación adjunta a un usuario, el usuario debe tener el permiso EXECUTE sobre la UDF para consultar la relación.
- La seguridad en el nivel de la fila puede limitar la optimización de las consultas. Recomendamos evaluar detenidamente el rendimiento de las consultas antes de implementar vistas protegidas por RLS en conjuntos de datos grandes.
- Las políticas de seguridad en el nivel de la fila que se aplican a las vistas de enlace tardío podrían incorporarse en las tablas federadas. Estas políticas de RLS pueden estar visibles en los registros de los motores de procesamiento externos.

## Limitaciones

A continuación, se indican las limitaciones al trabajar con políticas de RLS:

- Las políticas de RLS no se pueden asociar a tablas externas ni a muchos otros tipos de relaciones. Para obtener más información, consulte ATTACH RLS POLICY.
- Amazon Redshift admite las instrucciones SELECT para determinadas políticas de RLS con búsquedas que tienen uniones complejas, pero no admite las instrucciones UPDATE o DELETE. En los casos con instrucciones UPDATE o DELETE, Amazon Redshift devuelve el siguiente error:

```
ERROR: One of the RLS policies on target relation is not supported in UPDATE/DELETE.
```

- Siempre que se hace referencia a una función definida por el usuario (UDF) en una política de RLS de una relación adjunta a un usuario, el usuario debe tener el permiso EXECUTE sobre la UDF para consultar la relación.
- No se admiten las subconsultas correlacionadas. Amazon Redshift devuelve el siguiente error:

```
ERROR: RLS policy could not be rewritten.
```

- Amazon Redshift no admite el uso de recursos compartidos de datos con RLS. Si una relación no tiene el RLS desactivado para los recursos compartidos de datos, se produce un error en la consulta en el clúster de consumidores y se produce el siguiente error:

```
RLS-protected relation "rls_protected_table" cannot be accessed via datasharing
 query.
```

Puede desactivar el RLS para los recursos compartidos de datos mediante el comando ALTER TABLE con el parámetro ROW LEVEL SECURITY OFF FOR DATASHARES. Para obtener más información sobre el uso de ALTER TABLE para habilitar o deshabilitar RLS, vaya a ALTER TABLE.

- En las consultas entre bases de datos, Amazon Redshift bloquea las lecturas en las relaciones protegidas por RLS. Los usuarios con el permiso IGNORE RLS pueden acceder a la relación protegida mediante consultas entre bases de datos. Cuando un usuario sin el permiso IGNORE RLS accede a una relación protegida por RLS a través de una consulta entre bases de datos, aparece el siguiente error:

```
RLS-protected relation "rls_protected_table" cannot be accessed via cross-database
 query.
```

- ALTER RLS POLICY solo admite la modificación de una política de RLS mediante la cláusula USING (using_predicate_exp). No puede modificar una política de RLS con una cláusula WITH al ejecutar ALTER RLS POLICY.
- No puede consultar las relaciones que tienen activada la seguridad en el nivel de fila si los valores de alguna de las siguientes opciones de configuración no coinciden con el valor predeterminado de la sesión:
  - `enable_case_sensitive_super_attribute`
  - `enable_case_sensitive_identifier`
  - `downcase_delimited_identifier`

  Considere la posibilidad de restablecer las opciones de configuración de la sesión si intenta consultar una relación con la seguridad de nivel de fila activada y ve el mensaje “La relación protegida por RLS no admite la configuración de nivel de sesión porque la distinción entre mayúsculas y minúsculas es diferente de su valor predeterminado”.
- Cuando el clúster o espacio de nombres sin servidor aprovisionado tiene políticas de seguridad en el nivel de fila, los siguientes comandos están bloqueados para los usuarios habituales:

```
ALTER <current_user> SET enable_case_sensitive_super_attribute/
enable_case_sensitive_identifier/downcase_delimited_identifier
```

  Al crear políticas de RLS, le recomendamos que cambie las opciones de configuración predeterminadas para los usuarios normales para que coincidan con las opciones de configuración de la sesión en el momento en que se creó la política. Los superusuarios y los usuarios con el

privilegio ALTER USER pueden hacerlo mediante la configuración del grupo de parámetros o el comando ALTER USER. Para obtener información sobre grupos de parámetros, consulte Grupos de parámetros de Amazon Redshift en la Guía de administración de Amazon Redshift. Para obtener información sobre el comando ALTER USER, consulte ALTER USER.

- Las vistas y las vistas de enlace tardío con políticas de seguridad en el nivel de la fila y no se pueden sustituir por usuarios normales que utilicen el comando CREATE VIEW. Para reemplazar las vistas o LBV por políticas de RLS, primero debe separar las políticas de RLS asociadas a ellas, sustituir las vistas o LBV y volver a adjuntar las políticas. Los superusuarios y los usuarios que dispongan del `sys:secadmin permission` pueden utilizar CREATE VIEW en vistas o LBV con políticas de RLS sin tener que separar las políticas.
- Las vistas con políticas de seguridad en el nivel de la fila no pueden hacer referencia a las tablas y vistas del sistema.
- Una vista de enlace tardío a la que hace referencia una vista normal no puede estar protegida por RLS.
- No se puede acceder a las relaciones protegidas por RLS ni a los datos anidados de los lagos de datos en la misma consulta.

## Prácticas recomendadas para el rendimiento de RLS

A continuación, se enumeran las prácticas recomendadas para garantizar un mejor rendimiento de Amazon Redshift en las tablas protegidas por RLS.

### Seguridad de los operadores y las funciones

Al consultar tablas protegidas por RLS, el uso de determinados operadores o funciones puede conducir a un deterioro del rendimiento. Amazon Redshift clasifica a los operadores y las funciones como seguros o no seguros para consultar tablas protegidas por RLS. Una función u operador se clasifica como seguro para RLS cuando no tiene efectos secundarios observables en función de las entradas. Concretamente, una función u operador seguro para RLS no puede ser uno de los siguientes:

- Muestra un valor de entrada, o un valor que depende del valor de entrada, con o sin un mensaje de error.
- Falla o devuelve errores que dependen del valor de entrada.

Entre los operadores no seguros para RLS se incluyen los siguientes:

- Operadores aritméticos: +, -, /, *, %.
- Operadores de texto: LIKE y SIMILAR TO.
- Operadores cast.
- UDF.

Utilice la siguiente instrucción SELECT para comprobar la seguridad de los operadores y las funciones.

```
SELECT proname, proc_is_rls_safe(oid) FROM pg_proc;
```

Al planificar consultas en tablas protegidas por RLS, Amazon Redshift impone restricciones al orden de evaluación de los predicados de usuario que contienen operadores y funciones no seguros para RLS. Las consultas que hacen referencia a operadores o funciones no seguros para RLS pueden provocar un deterioro del rendimiento al consultar tablas protegidas por RLS. El rendimiento puede deteriorarse de manera significativa cuando Amazon Redshift no puede enviar predicados no seguros para RLS a los análisis de la tabla base para aprovechar las claves de clasificación. Para un mejor rendimiento, evite las consultas que utilizan predicados no seguros para RLS que aprovechan una clave de clasificación. Utilice las instrucciones EXPLAIN junto con el permiso del sistema EXPLAIN RLS para comprobar que Amazon Redshift puede enviar operadores y funciones.

## Almacenamiento en caché de los resultados

Para reducir el tiempo de ejecución de las consultas y mejorar el rendimiento del sistema, Amazon Redshift almacena en caché los resultados de determinados tipos de consultas en la memoria del nodo principal.

Amazon Redshift utiliza los resultados almacenados en caché para una consulta nueva que analiza tablas protegidas con RLS cuando se cumplen todas las condiciones de las tablas no protegidas y se cumplen todas las condiciones a continuación:

- Las tablas o las vistas de la política no se han modificado.
- La consulta no utiliza una función que debe evaluarse cada vez que se ejecuta, como GETDATE o CURRENT_USER.

Para obtener un mejor rendimiento, evite utilizar predicados de políticas que no cumplan con las condiciones anteriores.

Para obtener más información sobre el almacenamiento en caché de los resultados en Amazon Redshift, consulte Almacenamiento en caché de los resultados .

### Políticas complejas

Para obtener un mejor rendimiento, evite el uso de políticas complejas con subconsultas que unan varias tablas.

## Ejemplo completo de seguridad en el nivel de fila

A continuación, se muestra un ejemplo completo para ilustrar cómo un superusuario crea algunos usuarios y roles. Luego, un usuario con el rol secadmin crea, adjunta, desconecta y elimina políticas de RLS. En este ejemplo, se utiliza la base de datos de ejemplo tickit. Para obtener más información, consulte Cargar datos desde Amazon S3 en Amazon Redshift en la Guía de introducción a Amazon Redshift.

```
-- Create users and roles referenced in the policy statements.
CREATE ROLE analyst;
CREATE ROLE consumer;
CREATE ROLE dbadmin;
CREATE ROLE auditor;
CREATE USER bob WITH PASSWORD 'Name_is_bob_1';
CREATE USER alice WITH PASSWORD 'Name_is_alice_1';
CREATE USER joe WITH PASSWORD 'Name_is_joe_1';
CREATE USER molly WITH PASSWORD 'Name_is_molly_1';
CREATE USER bruce WITH PASSWORD 'Name_is_bruce_1';
GRANT ROLE sys:secadmin TO bob;
GRANT ROLE analyst TO alice;
GRANT ROLE consumer TO joe;
GRANT ROLE dbadmin TO molly;
GRANT ROLE auditor TO bruce;
GRANT ALL ON TABLE tickit_category_redshift TO PUBLIC;
GRANT ALL ON TABLE tickit_sales_redshift TO PUBLIC;
GRANT ALL ON TABLE tickit_event_redshift TO PUBLIC;

-- Create table and schema referenced in the policy statements.
CREATE SCHEMA target_schema;
GRANT ALL ON SCHEMA target_schema TO PUBLIC;
CREATE TABLE target_schema.target_event_table (LIKE tickit_event_redshift);
GRANT ALL ON TABLE target_schema.target_event_table TO PUBLIC;

-- Change session to analyst alice.
```

```
SET SESSION AUTHORIZATION alice;

-- Check the tuples visible to analyst alice.
-- Should contain all 3 categories.
SELECT catgroup, count(*)
FROM tickit_category_redshift
GROUP BY catgroup ORDER BY catgroup;

-- Change session to security administrator bob.
SET SESSION AUTHORIZATION bob;

CREATE RLS POLICY policy_concerts
WITH (catgroup VARCHAR(10))
USING (catgroup = 'Concerts');

SELECT poldb, polname, polalias, polatts, polqual, polenabled, polmodifiedby FROM
 svv_rls_policy WHERE poldb = CURRENT_DATABASE();

ATTACH RLS POLICY policy_concerts ON tickit_category_redshift TO ROLE analyst, ROLE
 dbadmin;

ALTER TABLE tickit_category_redshift ROW LEVEL SECURITY ON;

SELECT * FROM svv_rls_attached_policy;

-- Change session to analyst alice.
SET SESSION AUTHORIZATION alice;

-- Check that tuples with only `Concert` category will be visible to analyst alice.
SELECT catgroup, count(*)
FROM tickit_category_redshift
GROUP BY catgroup ORDER BY catgroup;

-- Change session to consumer joe.
SET SESSION AUTHORIZATION joe;

-- Although the policy is attached to a different role, no tuples will be
-- visible to consumer joe because the default deny all policy is applied.
SELECT catgroup, count(*)
FROM tickit_category_redshift
GROUP BY catgroup ORDER BY catgroup;

-- Change session to dbadmin molly.
SET SESSION AUTHORIZATION molly;
```

```
-- Check that tuples with only `Concert` category will be visible to dbadmin molly.
SELECT catgroup, count(*)
FROM tickit_category_redshift
GROUP BY catgroup ORDER BY catgroup;

-- Check that EXPLAIN output contains RLS SecureScan to prevent disclosure of
-- sensitive information such as RLS filters.
EXPLAIN SELECT catgroup, count(*) FROM tickit_category_redshift GROUP BY catgroup ORDER
 BY catgroup;

-- Change session to security administrator bob.
SET SESSION AUTHORIZATION bob;

-- Grant IGNORE RLS permission so that RLS policies do not get applicable to role
 dbadmin.
GRANT IGNORE RLS TO ROLE dbadmin;

-- Grant EXPLAIN RLS permission so that anyone in role auditor can view complete
 EXPLAIN output.
GRANT EXPLAIN RLS TO ROLE auditor;

-- Change session to dbadmin molly.
SET SESSION AUTHORIZATION molly;

-- Check that all tuples are visible to dbadmin molly because `IGNORE RLS` is granted
 to role dbadmin.
SELECT catgroup, count(*)
FROM tickit_category_redshift
GROUP BY catgroup ORDER BY catgroup;

-- Change session to auditor bruce.
SET SESSION AUTHORIZATION bruce;

-- Check explain plan is visible to auditor bruce because `EXPLAIN RLS` is granted to
 role auditor.
EXPLAIN SELECT catgroup, count(*) FROM tickit_category_redshift GROUP BY catgroup ORDER
 BY catgroup;

-- Change session to security administrator bob.
SET SESSION AUTHORIZATION bob;

DETACH RLS POLICY policy_concerts ON tickit_category_redshift FROM ROLE analyst, ROLE
 dbadmin;
```

```
-- Change session to analyst alice.
SET SESSION AUTHORIZATION alice;

-- Check that no tuples are visible to analyst alice.
-- Although the policy is detached, no tuples will be visible to analyst alice
-- because of default deny all policy is applied if the table has RLS on.
SELECT catgroup, count(*)
FROM tickit_category_redshift
GROUP BY catgroup ORDER BY catgroup;

-- Change session to security administrator bob.
SET SESSION AUTHORIZATION bob;

CREATE RLS POLICY policy_events
WITH (eventid INTEGER) AS ev
USING (
    ev.eventid IN (SELECT eventid FROM tickit_sales_redshift WHERE qtysold <3)
);

ATTACH RLS POLICY policy_events ON tickit_event_redshift TO ROLE analyst;
ATTACH RLS POLICY policy_events ON target_schema.target_event_table TO ROLE consumer;

RESET SESSION AUTHORIZATION;

-- Can not cannot alter type of dependent column.
ALTER TABLE target_schema.target_event_table ALTER COLUMN eventid TYPE float;
ALTER TABLE tickit_event_redshift ALTER COLUMN eventid TYPE float;
ALTER TABLE tickit_sales_redshift ALTER COLUMN eventid TYPE float;
ALTER TABLE tickit_sales_redshift ALTER COLUMN qtysold TYPE float;

-- Can not cannot rename dependent column.
ALTER TABLE target_schema.target_event_table RENAME COLUMN eventid TO renamed_eventid;
ALTER TABLE tickit_event_redshift RENAME COLUMN eventid TO renamed_eventid;
ALTER TABLE tickit_sales_redshift RENAME COLUMN eventid TO renamed_eventid;
ALTER TABLE tickit_sales_redshift RENAME COLUMN qtysold TO renamed_qtysold;

-- Can not drop dependent column.
ALTER TABLE target_schema.target_event_table DROP COLUMN eventid CASCADE;
ALTER TABLE tickit_event_redshift DROP COLUMN eventid CASCADE;
ALTER TABLE tickit_sales_redshift DROP COLUMN eventid CASCADE;
ALTER TABLE tickit_sales_redshift DROP COLUMN qtysold CASCADE;

-- Can not drop lookup table.
```

```
DROP TABLE tickit_sales_redshift CASCADE;

-- Change session to security administrator bob.
SET SESSION AUTHORIZATION bob;

DROP RLS POLICY policy_concerts;
DROP RLS POLICY IF EXISTS policy_events;

ALTER TABLE tickit_category_redshift ROW LEVEL SECURITY OFF;

RESET SESSION AUTHORIZATION;

-- Drop users and roles.
DROP USER bob;
DROP USER alice;
DROP USER joe;
DROP USER molly;
DROP USER bruce;
DROP ROLE analyst;
DROP ROLE consumer;
DROP ROLE auditor FORCE;
DROP ROLE dbadmin FORCE;
```

# Seguridad de los metadatos

De igual modo que la seguridad en el nivel de la fila de Amazon Redshift, la seguridad de los metadatos le proporciona un mayor control sobre los metadatos. Si la seguridad de los metadatos está habilitada para el clúster aprovisionado o el grupo de trabajo sin servidor, los usuarios pueden ver los metadatos de los objetos a los que tienen acceso de visualización. La seguridad de los metadatos le permite separar la visibilidad en función de sus necesidades. Por ejemplo, puede usar un único almacenamiento de datos para centralizar todo su almacenamiento de datos. Sin embargo, si almacena datos de varios sectores, administrar la seguridad puede convertirse en una tarea compleja. Con la seguridad de los metadatos habilitada, puede configurar su visibilidad. Los usuarios de un sector pueden tener más visibilidad sobre sus objetos, mientras que usted restringe el acceso de visualización a los usuarios de otro sector. La seguridad de los metadatos admite todos los tipos de objetos, como los esquemas, las tablas, las vistas, las vistas materializadas, los procedimientos almacenados, las funciones definidas por el usuario y los modelos de machine learning.

Los usuarios pueden ver los metadatos de los objetos en las siguientes circunstancias:

- Si se concede acceso a los objetos al usuario.

- Si el acceso al objeto se concede a un grupo o un rol al que pertenece el usuario.
- El objeto es público.
- El usuario es el propietario del objeto de base de datos.

Para habilitar la seguridad de los metadatos, utilice el comando ALTER SYSTEM. La siguiente es la sintaxis de cómo utilizar el comando ALTER SYSTEM con la seguridad de los metadatos.

```
ALTER SYSTEM SET metadata_security=[true|t|on|false|f|off];
```

Al habilitar la seguridad de los metadatos, todos los usuarios que tengan los permisos necesarios pueden ver los metadatos pertinentes de los objetos a los que tienen acceso. Si desea que solo determinados usuarios puedan ver la seguridad de los metadatos, conceda el permiso `ACCESS CATALOG` a un rol y, a continuación, asígnelo al usuario. Para obtener más información sobre el uso de los roles para controlar mejor la seguridad, consulte Control de acceso basado en roles.

En el siguiente ejemplo se muestra cómo conceder el permiso `ACCESS CATALOG` a un rol y, a continuación, asignar el rol a un usuario. Para obtener más información sobre cómo conceder permisos, consulte el comando GRANT.

```
CREATE ROLE sample_metadata_viewer;

GRANT ACCESS CATALOG TO ROLE sample_metadata_viewer;

GRANT ROLE sample_metadata_viewer to salesadmin;
```

Si prefiere usar roles ya definidos, los roles definidos por el sistema `operator`, `secadmin`, `dba` y `superuser` tienen todos los permisos necesarios para ver los metadatos de los objetos. De forma predeterminada, los superusuarios pueden ver el catálogo completo.

```
GRANT ROLE operator to sample_user;
```

Si usa roles para controlar la seguridad de los metadatos, tendrá acceso a todas las vistas y funciones del sistema que incluyan el control de acceso basado en roles. Por ejemplo, puede consultar la vista de SVV_ROLES para ver todos los roles. Para ver si un usuario es miembro de un rol o grupo, utilice la función USER_IS_MEMBER_OF. Para obtener una lista completa de las vistas SVV, consulte las vistas de metadatos SVV. Para obtener una lista de las funciones de información del sistema, consulte Funciones de información del sistema.

# Enmascaramiento de datos dinámico

Mediante el enmascaramiento dinámico de datos (DDM) en Amazon Redshift, puede proteger los datos confidenciales de su almacenamiento de datos. Puede manipular la forma en que Amazon Redshift muestra los datos confidenciales al usuario en el momento de la consulta, sin transformarlos en la base de datos. El acceso a los datos se controla mediante políticas de enmascaramiento que aplican reglas de ofuscación personalizadas a un usuario o rol determinados. De este modo, puede responder a los cambiantes requisitos de privacidad sin alterar los datos subyacentes ni editar las consultas SQL.

Las políticas de enmascaramiento dinámico de datos ocultan, ofuscan o seudonimizan los datos que coinciden con un formato determinado. Cuando se adjunta a una tabla, la expresión de enmascaramiento se aplica a una o varias de sus columnas. Puede modificar aún más las políticas de enmascaramiento para aplicarlas solo a determinados usuarios o a roles definidos por el usuario que puede crear con Control de acceso basado en roles (RBAC). Además, puede aplicar el enmascaramiento dinámico de datos en el nivel de celda mediante el uso de columnas condicionales al crear la política de enmascaramiento. Para obtener más información sobre el enmascaramiento condicional, consulte Enmascaramiento dinámico y condicional de datos.

Puede aplicar varias políticas de enmascaramiento con diferentes niveles de ofuscación a la misma columna de una tabla y asignarlas a diferentes roles. Para evitar conflictos cuando tiene diferentes roles con distintas políticas que se aplican a una columna, puede establecer prioridades para cada aplicación. De este modo, puede controlar a qué datos puede acceder un determinado usuario o rol. Las políticas de enmascaramiento dinámico de datos pueden elaborar parcial o totalmente los datos, o crear un hash de ellos utilizando funciones definidas por el usuario escritas en SQL o Python, o con AWS Lambda. Al enmascarar los datos mediante hashes, puede aplicar combinaciones en estos datos sin acceder a información potencialmente confidencial.

## Comandos SQL para la administración de políticas de enmascaramiento de datos dinámico

Puede realizar las siguientes acciones para crear, adjuntar, separar y eliminar políticas de enmascaramiento de datos dinámico:

- Para crear una política de enmascaramiento dinámico de datos, utilice el comando CREATE MASKING POLICY.

A continuación, se ofrece un ejemplo de cómo crear una política de enmascaramiento mediante una función de hash SHA-2.

```
CREATE MASKING POLICY hash_credit
WITH (credit_card varchar(256))
USING (sha2(credit_card + 'testSalt', 256));
```

- Para modificar una política DDM existente, utilice el comando ALTER MASKING POLICY.

  A continuación, se ofrece un ejemplo de modificación de una política de enmascaramiento existente.

```
ALTER MASKING POLICY hash_credit
USING (sha2(credit_card + 'otherTestSalt', 256));
```

- Para adjuntar una política de enmascaramiento dinámico de datos a una tabla para uno o varios usuarios o roles, utilice el comando ATTACH MASKING POLICY.

  A continuación, se ofrece un ejemplo de una política de enmascaramiento en un par de columna-rol.

```
 ATTACH MASKING POLICY hash_credit
ON credit_cards (credit_card)
TO ROLE science_role
PRIORITY 30;
```

  La cláusula PRIORITY determina qué política de enmascaramiento se aplica a una sesión de usuario cuando se adjuntan varias políticas a la misma columna. Por ejemplo, si el usuario del ejemplo anterior tiene otra política de enmascaramiento adjunta a la misma columna de tarjetas de crédito con una prioridad de 20, la política de science_role es la que se aplica, ya que tiene la prioridad más alta, es decir, 30.

- Para desconectar una política de enmascaramiento dinámico de datos de una tabla de uno o varios usuarios o roles, utilice el comando DETACH MASKING POLICY.

  A continuación, se ofrece un ejemplo de desconexión de una política de enmascaramiento de un par de columna-rol.

```
DETACH MASKING POLICY hash_credit
ON credit_cards(credit_card)
```

```
FROM ROLE science_role;
```

- Para eliminar una política de enmascaramiento dinámico de datos de todas las bases de datos, utilice el comando DROP MASKING POLICY.

  A continuación, se ofrece un ejemplo de cómo eliminar una política de enmascaramiento de todas las bases de datos.

```
DROP MASKING POLICY hash_credit;
```

## Jerarquía de políticas de enmascaramiento de datos dinámico

Cuando asocie varias políticas de enmascaramiento, tenga en cuenta lo siguiente:

- Se pueden adjuntar varias políticas de enmascaramiento a una sola columna.
- Cuando se aplican varias políticas de enmascaramiento a una consulta, se aplica la política de mayor prioridad adjuntada a cada columna. Considere el siguiente ejemplo.

```
ATTACH MASKING POLICY partial_hash
ON credit_cards(address, credit_card)
TO ROLE analytics_role
PRIORITY 20;

ATTACH MASKING POLICY full_hash
ON credit_cards(credit_card, ssn)
TO ROLE auditor_role
PRIORITY 30;

SELECT address, credit_card, ssn
FROM credit_cards;
```

  Al ejecutar la instrucción SELECT, un usuario con los roles de analista y auditor ve la columna de direcciones con la política de enmascaramiento `partial_hash` aplicada. También ve las columnas de tarjetas de crédito y SSN con la política de enmascaramiento `full_hash` aplicada, porque la política `full_hash` tiene una mayor prioridad en la columna de tarjetas de crédito.
- Si no especifica una prioridad al asociar una política de enmascaramiento, la prioridad predeterminada es 0.
- No puede adjuntar dos políticas a la misma columna con la misma prioridad.

- No puede adjuntar dos políticas a la misma combinación de usuario y columna o rol y columna.
- Cuando se aplican varias políticas de enmascaramiento a lo largo de la misma ruta SUPER mientras estén asociadas al mismo usuario o rol, solo surtirá efecto la vinculación de mayor prioridad. Considere los siguientes ejemplos:

  El primer ejemplo muestra dos políticas de enmascaramiento asociadas en la misma ruta donde se aplica la política de mayor prioridad.

```
ATTACH MASKING POLICY hide_name
ON employees(col_person.name)
TO PUBLIC
PRIORITY 20;

ATTACH MASKING POLICY hide_last_name
ON employees(col_person.name.last)
TO PUBLIC
PRIORITY 30;

--Only the hide_last_name policy takes effect.
SELECT employees.col_person.name FROM employees;
```

  El segundo ejemplo muestra dos políticas de enmascaramiento asociadas a rutas diferentes en el mismo objeto SUPER, sin ningún conflicto entre las políticas. Ambos adjuntos se aplicarán de forma simultánea.

```
ATTACH MASKING POLICY hide_first_name
ON employees(col_person.name.first)
TO PUBLIC
PRIORITY 20;

ATTACH MASKING POLICY hide_last_name
ON employees(col_person.name.last)
TO PUBLIC
PRIORITY 20;

--Both col_person.name.first and col_person.name.last are masked.
SELECT employees.col_person.name FROM employees;
```

Para confirmar qué política de enmascaramiento se aplica a una determinada combinación de usuario y columna o rol y columna, los usuarios con el rol `sys:secadmin` pueden buscar el par de

columna y rol o columna y usuario en la vista del sistema SVV_ATTACHED_MASKING_POLICY. Para obtener más información, consulte Vistas del sistema de enmascaramiento de datos dinámico.

## Uso del enmascaramiento dinámico de datos con rutas de tipos de datos SUPER

Amazon Redshift permite asociar políticas de enmascaramiento dinámico de datos a las rutas de las columnas de tipo SUPER. Para obtener más información acerca del tipo de datos SUPER, consulte Datos semiestructurados en Amazon Redshift.

Al asociar políticas de enmascaramiento a las rutas de las columnas de tipo SUPER, tenga en cuenta lo siguiente.

- Al asociar una política de enmascaramiento a una ruta de una columna, esa columna debe definirse como el tipo de datos SUPER. Solo puede aplicar políticas de enmascaramiento a los valores escalares de la ruta SUPER. No puede aplicar políticas de enmascaramiento a estructuras o matrices complejas.
- Puede aplicar diferentes políticas de enmascaramiento a varios valores escalares de una sola columna SUPER, siempre que las rutas SUPER no entren en conflicto. Por ejemplo, las rutas SUPER `a.b` y `a.b.c` entran en conflicto porque están en la misma ruta. Además, `a.b` es el principal de `a.b.c`. Las rutas SUPER `a.b.c` y `a.b.d` no entran en conflicto.
- Amazon Redshift no puede comprobar si las rutas asociadas a una política de enmascaramiento existen en los datos y son del tipo esperado hasta que la política se aplique en el tiempo de ejecución de las consultas de los usuarios. Por ejemplo, si asocia una política de enmascaramiento que enmascara valores TEXT a una ruta SUPER que contiene un valor INT, Amazon Redshift intenta convertir el tipo de valor en la ruta.

  En estas situaciones, el comportamiento de Amazon Redshift en el tiempo de ejecución depende de los ajustes de la configuración de la consulta de objetos SUPER. De forma predeterminada, Amazon Redshift está en modo laxo y resolverá las rutas que falten y las cadenas no válidas como `NULL` en la ruta SUPER dada. Para obtener más información acerca de los valores de configuración relacionados con SUPER, consulte Configuraciones SUPER.
- SUPER es un tipo sin esquema, lo que significa que Amazon Redshift no puede confirmar la existencia del valor en una ruta SUPER determinada. Si asocia una política de enmascaramiento a una ruta SUPER que no existe y Amazon Redshift está en modo laxo, Amazon Redshift resolverá la ruta hacia un valor `NULL`. Le recomendamos que tenga en cuenta el formato esperado de los objetos SUPER y la probabilidad de que tengan atributos inesperados al asociar políticas de

enmascaramiento a las rutas de las columnas SUPER. Si cree que puede haber un esquema inesperado en su columna SUPER, considere asociar sus políticas de enmascaramiento directamente a la columna SUPER. Puede utilizar las funciones de información de tipo SUPER para comprobar los atributos y los tipos, y utilizar `OBJECT_TRANSFORM` para enmascarar los valores. Para obtener más información acerca de las funciones de información de tipo SUPER, consulte Funciones de información acerca del tipo SUPER.

## Ejemplos

Asociación de políticas de enmascaramiento a las rutas SUPER

En el siguiente ejemplo, se asocian varias políticas de enmascaramiento a varias rutas de tipo SUPER en una columna.

```
CREATE TABLE employees (
    col_person SUPER
);

INSERT INTO employees
VALUES
    (
        json_parse('
            {
                "name": {
                    "first": "John",
                    "last": "Doe"
                },
                "age": 25,
                "ssn": "111-22-3333",
                "company": "Company Inc."
            }
        ')
    ),
    (
        json_parse('
            {
                "name": {
                    "first": "Jane",
                    "last": "Appleseed"
                },
                "age": 34,
                "ssn": "444-55-7777",
```

```
                "company": "Organization Org."
            }
        ')
    )
;
GRANT ALL ON ALL TABLES IN SCHEMA "public" TO PUBLIC;

-- Create the masking policies.

-- This policy converts the given name to all uppercase letters.
CREATE MASKING POLICY mask_first_name
WITH(first_name TEXT)
USING ( UPPER(first_name) );

-- This policy replaces the given name with the fixed string 'XXXX'.
CREATE MASKING POLICY mask_last_name
WITH(last_name TEXT)
USING ( 'XXXX'::TEXT );

-- This policy rounds down the given age to the nearest 10.
CREATE MASKING POLICY mask_age
WITH(age INT)
USING ( (FLOOR(age::FLOAT / 10) * 10)::INT );

-- This policy converts the first five digits of the given SSN to 'XXX-XX'.
CREATE MASKING POLICY mask_ssn
WITH(ssn TEXT)
USING ( 'XXX-XX-'::TEXT || SUBSTRING(ssn::TEXT FROM 8 FOR 4) );

-- Attach the masking policies to the employees table.
ATTACH MASKING POLICY mask_first_name
ON employees(col_person.name.first)
TO PUBLIC;

ATTACH MASKING POLICY mask_last_name
ON employees(col_person.name.last)
TO PUBLIC;

ATTACH MASKING POLICY mask_age
ON employees(col_person.age)
TO PUBLIC;

ATTACH MASKING POLICY mask_ssn
ON employees(col_person.ssn)
```

```
TO PUBLIC;

-- Verify that your masking policies are attached.
SELECT
    policy_name,
    TABLE_NAME,
    priority,
    input_columns,
    output_columns
FROM
    svv_attached_masking_policy;

   policy_name   | table_name | priority |           input_columns           |
 output_columns
-----------------+------------+----------+-----------------------------------
+-----------------------------------
 mask_age        | employees  |        0 | ["col_person.\"age\""]            |
 ["col_person.\"age\""]
 mask_first_name | employees  |        0 | ["col_person.\"name\".\"first\""] |
 ["col_person.\"name\".\"first\""]
 mask_last_name  | employees  |        0 | ["col_person.\"name\".\"last\""]  |
 ["col_person.\"name\".\"last\""]
 mask_ssn        | employees  |        0 | ["col_person.\"ssn\""]            |
 ["col_person.\"ssn\""]
(4 rows)

-- Observe the masking policies taking effect.
SELECT col_person FROM employees ORDER BY col_person.age;

-- This result is formatted for ease of reading.
         col_person
-------------------------------
{
    "name": {
        "first": "JOHN",
        "last": "XXXX"
    },
    "age": 20,
    "ssn": "XXX-XX-3333",
    "company": "Company Inc."
}
{
    "name": {
        "first": "JANE",
```

```
        "last": "XXXX"
    },
    "age": 30,
    "ssn": "XXX-XX-7777",
    "company": "Organization Org."
}
```

A continuación se muestran algunos ejemplos de asociaciones no válidas de políticas de enmascaramiento a rutas SUPER.

```
-- This attachment fails because there is already a policy
-- with equal priority attached to employees.name.last, which is
-- on the same SUPER path as employees.name.
ATTACH MASKING POLICY mask_ssn
ON employees(col_person.name)
TO PUBLIC;
ERROR:  DDM policy "mask_last_name" is already attached on relation "employees" column
 "col_person."name"."last"" with same priority

-- Create a masking policy that masks DATETIME objects.
CREATE MASKING POLICY mask_date
WITH(INPUT DATETIME)
USING ( INPUT );

-- This attachment fails because SUPER type columns can't contain DATETIME objects.
ATTACH MASKING POLICY mask_date
ON employees(col_person.company)
TO PUBLIC;
ERROR:  cannot attach masking policy for output of type "timestamp without time zone"
 to column "col_person."company"" of type "super
```

A continuación, se ofrece un ejemplo de una política de enmascaramiento a una ruta SUPER que no existe. De forma predeterminada, Amazon Redshift resolverá la ruta hacia `NULL`.

```
ATTACH MASKING POLICY mask_first_name
ON employees(col_person.not_exists)
TO PUBLIC;

SELECT col_person FROM employees LIMIT 1;

-- This result is formatted for ease of reading.
         col_person
```

```
-----------------------------------
{
    "name": {
        "first": "JOHN",
        "last": "XXXX"
    },
    "age": 20,
    "ssn": "XXX-XX-3333",
    "company": "Company Inc.",
    "not_exists": null
}
```

## Enmascaramiento dinámico y condicional de datos

Puede enmascarar datos en el nivel de celda mediante la creación de políticas de enmascaramiento con expresiones condicionales en la expresión de enmascaramiento. Por ejemplo, puede crear una política de enmascaramiento que aplique diferentes máscaras a un valor, en función del valor de otra columna de esa fila.

A continuación, se muestra un ejemplo de uso del enmascaramiento condicional de datos para crear y adjuntar una política de enmascaramiento que elabora parcialmente los números de tarjeta de crédito implicados en un fraude, al mismo tiempo que oculta completamente todos los demás números de tarjeta de crédito. Debe ser superusuario o tener el rol `sys:secadmin` para ejecutar este ejemplo.

```
--Create an analyst role.
CREATE ROLE analyst;

--Create a credit card table. The table contains an is_fraud boolean column,
--which is TRUE if the credit card number in that row was involved in a fraudulent
 transaction.
CREATE TABLE credit_cards (id INT, is_fraud BOOLEAN, credit_card_number VARCHAR(16));

--Create a function that partially redacts credit card numbers.
CREATE FUNCTION REDACT_CREDIT_CARD (credit_card VARCHAR(16))
RETURNS VARCHAR(16) IMMUTABLE
AS $$
    import re
    regexp = re.compile("^([0-9]{6})[0-9]{5,6}([0-9]{4})")

    match = regexp.search(credit_card)
    if match != None:
```

```
        first = match.group(1)
        last = match.group(2)
    else:
        first = "000000"
        last = "0000"

    return "{}XXXXX{}".format(first, last)
$$ LANGUAGE plpythonu;

--Create a masking policy that partially redacts credit card numbers if the is_fraud
 value for that row is TRUE,
--and otherwise blanks out the credit card number completely.
CREATE MASKING POLICY card_number_conditional_mask
    WITH (fraudulent BOOLEAN, pan varchar(16))
    USING (CASE WHEN fraudulent THEN REDACT_CREDIT_CARD(pan)
                ELSE Null
           END);

--Attach the masking policy to the credit_cards/analyst table/role pair.
ATTACH MASKING POLICY card_number_conditional_mask ON credit_cards (credit_card_number)
 USING (is_fraud, credit_card_number)
 TO ROLE analyst PRIORITY 100;
```

## Vistas del sistema de enmascaramiento de datos dinámico

Los superusuarios, los usuarios con el rol `sys:operator` y los usuarios con el permiso ACCESS SYSTEM TABLE pueden acceder a las siguientes vistas del sistema relacionadas con DDM.

- SVV_MASKING_POLICY

  Utilice SVV_MASKING_POLICY para ver todas las políticas de enmascaramiento creadas en el clúster o grupo de trabajo.

- SVV_ATTACHED_MASKING_POLICY

  Utilice SVV_ATTACHED_MASKING_POLICY para ver todas las relaciones y los usuarios o los roles con políticas adjuntas a la base de datos conectada actualmente.

- SYS_APPLIED_MASKING_POLICY_LOG

  Utilice SYS_APPLIED_MASKING_POLICY_LOG para rastrear la aplicación de políticas de enmascaramiento en consultas que hacen referencia a relaciones protegidas por DDM.

A continuación, se muestran algunos ejemplos de la información que puede encontrar utilizando las vistas del sistema.

```
--Select all policies associated with specific users, as opposed to roles
SELECT policy_name,
       schema_name,
       table_name,
       grantee
FROM svv_attached_masking_policy
WHERE grantee_type = 'user';

--Select all policies attached to a specific user
SELECT policy_name,
       schema_name,
       table_name,
       grantee
FROM svv_attached_masking_policy
WHERE grantee = 'target_grantee_name'

--Select all policies attached to a given table
SELECT policy_name,
       schema_name,
       table_name,
       grantee
FROM svv_attached_masking_policy
WHERE table_name = 'target_table_name'
      AND schema_name = 'target_schema_name';

--Select the highest priority policy attachment for a given role
SELECT samp.policy_name,
       samp.priority,
       samp.grantee,
       smp.policy_expression
FROM svv_masking_policy AS smp
JOIN svv_attached_masking_policy AS samp
    ON samp.policy_name = smp.policy_name
WHERE
    samp.grantee_type = 'role' AND
    samp.policy_name = mask_get_policy_for_role_on_column(
        'target_schema_name',
        'target_table_name',
        'target_column_name',
        'target_role_name')
```

```
ORDER BY samp.priority desc
LIMIT 1;

--See which policy a specific user will see on a specific column in a given relation
SELECT samp.policy_name,
       samp.priority,
       samp.grantee,
       smp.policy_expression
FROM svv_masking_policy AS smp
JOIN svv_attached_masking_policy AS samp
    ON samp.policy_name = smp.policy_name
WHERE
    samp.grantee_type = 'role' AND
    samp.policy_name = mask_get_policy_for_user_on_column(
        'target_schema_name',
        'target_table_name',
        'target_column_name',
        'target_user_name')
ORDER BY samp.priority desc;

 --Select all policies attached to a given relation.
SELECT policy_name,
schema_name,
relation_name,
database_name
FROM sys_applied_masking_policy_log
WHERE relation_name = 'relation_name'
AND schema_name = 'schema_name';
```

## Consideraciones al utilizar el enmascaramiento dinámico de datos

Cuando utilice el enmascaramiento dinámico de datos, tenga en cuenta lo siguiente:

- Al consultar objetos creados a partir de tablas, como las vistas, los usuarios verán los resultados según sus propias políticas de enmascaramiento, no las políticas del usuario que creó los objetos. Por ejemplo, un usuario con el rol de analista que consulte una vista creada por un administrador de seguridad (secadmin) vería resultados con políticas de enmascaramiento adjuntas al rol de analista.
- Para evitar que el comando EXPLAIN exponga potencialmente filtros confidenciales de políticas de enmascaramiento, solo los usuarios con el permiso SYS_EXPLAIN_DDM pueden ver las políticas

de enmascaramiento aplicadas en los resultados de EXPLAIN. Los usuarios no tienen el permiso SYS_EXPLAIN_DDM de forma predeterminada.

A continuación, se muestra la sintaxis para conceder el permiso a un rol.

```
GRANT EXPLAIN MASKING TO ROLE rolename
```

Para obtener más información sobre el comando EXPLAIN, consulte EXPLAIN.

- Los usuarios con distintos roles pueden ver resultados diferentes en función de las condiciones de filtrado o de combinación utilizadas. Por ejemplo, se producirá un error en la ejecución de un comando SELECT en una tabla mediante el valor de una columna específica si el usuario que ejecuta el comando tiene aplicada una política de enmascaramiento que ofusca esa columna.
- Las políticas de enmascaramiento dinámico de datos se deben aplicar antes de cualquier operación de predicado o proyección. Las políticas de enmascaramiento pueden incluir lo siguiente:
  - Operaciones constantes de bajo costo como convertir un valor en nulo
  - Operaciones de costo moderado, por ejemplo, creación de hash HMAC
  - Operaciones de costo elevado, como las llamadas a funciones Lambda externas definidas por el usuario

  Por lo tanto, le recomendamos que utilice expresiones de enmascaramiento simples siempre que sea posible.
- Puede utilizar políticas de enmascaramiento dinámico de datos para roles con políticas de seguridad a nivel de fila, pero tenga en cuenta que las políticas de RLS se aplican antes que las de enmascaramiento dinámico de datos. Una expresión de enmascaramiento dinámico de datos no podrá leer una fila protegida por RLS. Para obtener más información sobre RLS, consulte Seguridad de nivel básico.
- Cuando utilice el comando COPY para copiar de parquet a tablas de destino protegidas, deberá especificar explícitamente las columnas en la instrucción COPY. Para obtener más información sobre la asignación de columnas con COPY, consulte Opciones de mapeo de columnas.
- Las políticas de DDM no se pueden asociar a las siguientes relaciones:
  - Tablas y catálogos del sistema
  - tablas externas
  - Tablas de recursos compartidos de datos
  - Vistas materializadas

  - Relaciones entre bases de datos
  - Tablas temporales
  - Consultas correlacionadas
- Las políticas de DDM pueden incluir tablas de consulta. Las tablas de consulta pueden estar presentes en la cláusula USING. Los siguientes tipos de relaciones no se pueden usar como tablas de consulta:
  - Tablas y catálogos del sistema
  - tablas externas
  - Tablas de recursos compartidos de datos
  - Vistas, vistas materializadas y vistas de enlace tardío
  - Relaciones entre bases de datos
  - Tablas temporales
  - Consultas correlacionadas

  A continuación, se ofrece un ejemplo de una política de enmascaramiento a una tabla de búsqueda.

```
--Create a masking policy referencing a lookup table
CREATE MASKING POLICY lookup_mask_credit_card WITH (credit_card TEXT) USING (
  CASE
    WHEN
      credit_card IN (SELECT credit_card_lookup FROM credit_cards_lookup)
    THEN '000000XXXX0000'
    ELSE REDACT_CREDIT_CARD(credit_card)
    END
  );

--Provides access to the lookup table via a policy attached to a role
GRANT SELECT ON TABLE credit_cards_lookup TO MASKING POLICY lookup_mask_credit_card;
```

- No puede adjuntar una política de enmascaramiento que produzca un resultado incompatible con el tipo y el tamaño de la columna de destino. Por ejemplo, no puede adjuntar una política de enmascaramiento que produzca una cadena de 12 caracteres de longitud para una columna VARCHAR(10). Amazon Redshift admite las siguientes excepciones:
  - Una política de enmascaramiento con el tipo de entrada INTN puede adjuntarse a una política con un tamaño INTM siempre que M < N. Por ejemplo, una política de entrada BIGINT (INT8) puede adjuntarse a una columna smallint (INT4).

  - Una política de enmascaramiento con el tipo de entrada NUMERIC o DECIMAL siempre puede adjuntarse a una columna FLOAT.
- Las políticas de enmascaramiento dinámico de datos no pueden utilizarse con el uso compartido de datos. Si el productor de datos del recurso compartido de datos asocia una política de enmascaramiento dinámico de datos a una tabla del recurso compartido de datos, dicha tabla se convierte en inaccesible para los usuarios del consumidor de datos que intentan consultarla. Las tablas con políticas de enmascaramiento dinámico de datos asociadas no pueden agregarse a un recurso compartido de datos.
- Amazon Redshift no admite el uso compartido de datos con DDM. Si una relación tiene activado el DDM para los recursos compartidos de datos, al intentar agregar la relación a un recurso compartido de datos se producirá el siguiente error en el clúster o el espacio de nombres del productor:

  ```
  <ddm_protected_relation> or a relation dependent on it is protected by a masking
   policy and cannot be added to a datashare
  ```

  Si adjunta una política de enmascaramiento a una relación del productor y la relación ya está incluida en un recurso compartido de datos, al intentar consultar la relación en el lado del consumidor, se producirá el siguiente error:

  ```
  cross-cluster query of the masked relation <ddm_protected_relation> is not supported.
  ```

  Puede desactivar el DDM para los recursos compartidos de datos mediante el comando ALTER TABLE con el parámetro MASKING OFF FOR DATASHARES. Para obtener más información, consulte ALTER TABLE.
- No puede consultar las relaciones que tienen asociadas políticas de enmascaramiento dinámico de datos si los valores de alguna de las siguientes opciones de configuración no coinciden con el valor predeterminado de la sesión:
  - `enable_case_sensitive_super_attribute`
  - `enable_case_sensitive_identifier`
  - `downcase_delimited_identifier`

  Considere la posibilidad de restablecer las opciones de configuración de la sesión si intenta consultar una relación con una política de enmascaramiento dinámico de datos asociada y ve el mensaje “La relación protegida de enmascaramiento dinámico de datos no admite la configuración

de nivel de sesión porque la distinción entre mayúsculas y minúsculas es diferente de su valor predeterminado".

- Cuando el clúster o espacio de nombres sin servidor aprovisionado tiene alguna política de enmascaramiento de datos dinámica, los siguientes comandos están bloqueados para los usuarios habituales:

```
ALTER <current_user> SET enable_case_sensitive_super_attribute/
enable_case_sensitive_identifier/downcase_delimited_identifier
```

  Al crear políticas de enmascaramiento dinámico de datos, le recomendamos que cambie las opciones de configuración predeterminadas para los usuarios habituales para que coincidan con las opciones de configuración de la sesión en el momento en que se creó la política. Los superusuarios y los usuarios con el privilegio ALTER USER pueden hacerlo mediante la configuración del grupo de parámetros o el comando ALTER USER. Para obtener información sobre grupos de parámetros, consulte Grupos de parámetros de Amazon Redshift en la Guía de administración de Amazon Redshift. Para obtener información sobre el comando ALTER USER, consulte ALTER USER.
- Los usuarios habituales no pueden reemplazar las vistas y las vistas de enlace tardío con políticas de DDM asociadas que utilicen el comando CREATE VIEW. Para reemplazar las vistas o LBV por políticas de RLS, primero debe separar las políticas de RLS asociadas a ellas, sustituir las vistas o LBV y volver a adjuntar las políticas. Los superusuarios y los usuarios que dispongan del permiso `sys:secadmin` pueden utilizar CREATE VIEW en vistas o LBV con políticas de DDM sin tener que separar las políticas.
- Las vistas con políticas de DDM asociadas no pueden hacer referencia a las tablas y vistas del sistema. Las vistas de enlace tardío pueden hacer referencia a tablas y vistas del sistema.
- Las vistas de enlace tardío con políticas de DDM asociadas no pueden hacer referencia a los datos anidados de los lagos de datos, como los documentos JSON.
- Las vistas de enlace tardío no pueden tener políticas de DDM asociadas si alguna vista hace referencia a esa vista de enlace tardío.
- Las políticas de DDM asociadas a las vistas de enlace tardío se asocian por nombre de columna. En el momento de la consulta, Amazon Redshift valida que todas las políticas de enmascaramiento asociadas a la vista de enlace tardío se hayan aplicado correctamente y que el tipo de columna de salida de la vista de enlace tardío coincida con los tipos de las políticas de enmascaramiento asociadas. Si se produce un error en la validación, Amazon Redshift devuelve un error para la consulta.

- Puede utilizar variables de contexto de sesión personalizadas al crear políticas de DDM. En el ejemplo siguiente se establecen variables de contexto para una política de DDM.

```
-- Set a customized context variable.
SELECT set_config('app.city', 'XXXX', FALSE);

-- Create a MASKING policy using current_setting() to get the value of a customized
 context variable.
CREATE MASKING POLICY city_mask
WITH (city VARCHAR(30))
USING (current_setting('app.city')::VARCHAR(30));

-- Attach the policy on the target table to one or more roles.
ATTACH MASKING POLICY city_mask
ON tickit_users_redshift(city)
TO ROLE analyst, ROLE dbadmin;
```

Para obtener más detalles sobre cómo configurar y recuperar variables de contexto de sesión personalizadas, vaya a SET, SET_CONFIG, SHOW, CURRENT_SETTING y RESET. Para obtener más información sobre la modificación de la configuración del servidor en general, vaya a Modificación de la configuración del servidor.

> **Important**
>
> Cuando se utilizan variables de contexto de sesión en las políticas de DDM, la política de seguridad depende del usuario o rol que la invoca. Tenga cuidado de evitar vulnerabilidades de seguridad al utilizar variables de contexto de sesión en las políticas de DDM.

## Ejemplo completo de enmascaramiento de datos dinámico

A continuación, se presenta un ejemplo completo en el que se muestra cómo puede crear y adjuntar políticas de enmascaramiento a una columna. Estas políticas permiten a los usuarios acceder a una columna y ver diferentes valores, según el grado de ofuscación de las políticas asociadas a sus roles. Debe ser superusuario o tener el rol `sys:secadmin` para ejecutar este ejemplo.

### Creación de una política de enmascaramiento

En primer lugar, cree una tabla y rellénela con los valores de tarjeta de crédito.

```
--create the table
CREATE TABLE credit_cards (
  customer_id INT,
  credit_card TEXT
);

--populate the table with sample values
INSERT INTO credit_cards
VALUES
  (100, '4532993817514842'),
  (100, '4716002041425888'),
  (102, '5243112427642649'),
  (102, '6011720771834675'),
  (102, '6011378662059710'),
  (103, '373611968625635')
;

--run GRANT to grant permission to use the SELECT statement on the table
GRANT SELECT ON credit_cards TO PUBLIC;

--create two users
CREATE USER regular_user WITH PASSWORD '1234Test!';

CREATE USER analytics_user WITH PASSWORD '1234Test!';

--create the analytics_role role and grant it to analytics_user
--regular_user does not have a role
CREATE ROLE analytics_role;

GRANT ROLE analytics_role TO analytics_user;
```

A continuación, cree una política de enmascaramiento para aplicarla al rol de análisis.

```
--create a masking policy that fully masks the credit card number
CREATE MASKING POLICY mask_credit_card_full
WITH (credit_card VARCHAR(256))
USING ('000000XXXX0000'::TEXT);

--create a user-defined function that partially obfuscates credit card data
CREATE FUNCTION REDACT_CREDIT_CARD (credit_card TEXT)
RETURNS TEXT IMMUTABLE
AS $$
    import re
```

```
    regexp = re.compile("^([0-9]{6})[0-9]{5,6}([0-9]{4})")

    match = regexp.search(credit_card)
    if match != None:
        first = match.group(1)
        last = match.group(2)
    else:
        first = "000000"
        last = "0000"

    return "{}XXXXX{}".format(first, last)
$$ LANGUAGE plpythonu;

--create a masking policy that applies the REDACT_CREDIT_CARD function
CREATE MASKING POLICY mask_credit_card_partial
WITH (credit_card VARCHAR(256))
USING (REDACT_CREDIT_CARD(credit_card));

--confirm the masking policies using the associated system views
SELECT * FROM svv_masking_policy;

SELECT * FROM svv_attached_masking_policy;
```

## Asociación de una política de enmascaramiento

Adjunte las políticas de enmascaramiento a la tabla de tarjetas de crédito.

```
--attach mask_credit_card_full to the credit card table as the default policy
--all users will see this masking policy unless a higher priority masking policy is
 attached to them or their role
ATTACH MASKING POLICY mask_credit_card_full
ON credit_cards(credit_card)
TO PUBLIC;

--attach mask_credit_card_partial to the analytics role
--users with the analytics role can see partial credit card information
ATTACH MASKING POLICY mask_credit_card_partial
ON credit_cards(credit_card)
TO ROLE analytics_role
PRIORITY 10;

--confirm the masking policies are applied to the table and role in the associated
 system view
```

```
SELECT * FROM svv_attached_masking_policy;

--confirm the full masking policy is in place for normal users by selecting from the
 credit card table as regular_user
SET SESSION AUTHORIZATION regular_user;

SELECT * FROM credit_cards;

--confirm the partial masking policy is in place for users with the analytics role by
 selecting from the credit card table as analytics_user
SET SESSION AUTHORIZATION analytics_user;

SELECT * FROM credit_cards;
```

## Creación de una política de enmascaramiento

En la siguiente sección, se muestra cómo modificar una política de enmascaramiento de datos dinámico.

```
--reset session authorization to the default
RESET SESSION AUTHORIZATION;

--alter the mask_credit_card_full policy
ALTER MASKING POLICY mask_credit_card_full
USING ('00000000000000'::TEXT);

--confirm the full masking policy is in place after altering the policy, and that
 results are altered from '000000XXXX0000' to '00000000000000'
SELECT * FROM credit_cards;
```

## Desconexión y eliminación de una política de enmascaramiento

En la siguiente sección se muestra cómo desconectar y eliminar las políticas de enmascaramiento mediante la eliminación de todas las políticas de enmascaramiento de datos dinámicos de la tabla.

```
--reset session authorization to the default
RESET SESSION AUTHORIZATION;

--detach both masking policies from the credit_cards table
DETACH MASKING POLICY mask_credit_card_full
ON credit_cards(credit_card)
FROM PUBLIC;
```

```
DETACH MASKING POLICY mask_credit_card_partial
ON credit_cards(credit_card)
FROM ROLE analytics_role;

--drop both masking policies
DROP MASKING POLICY mask_credit_card_full;

DROP MASKING POLICY mask_credit_card_partial;
```

# Permisos acotados

Los permisos limitados le permiten conceder permisos a un usuario o rol en todos los objetos de un tipo dentro de una base de datos o un esquema. Los usuarios y roles con permisos limitados tienen los permisos especificados en todos los objetos actuales y futuros de la base de datos o del esquema.

Puede ver el alcance de los permisos limitados en el nivel de base de datos en SVV_DATABASE_PRIVILEGES. Puede ver el alcance de los permisos limitados en el nivel de esquema en SVV_SCHEMA_PRIVILEGES.

Para obtener más información sobre la aplicación de permisos acotados, consulte GRANT y REVOKE.

## Consideraciones para utilizar permisos acotados

Cuando utilice permisos acotados, tenga en cuenta lo siguiente:

- Puede utilizar permisos acotados para conceder o revocar permisos en el ámbito de una base de datos o un esquema a o desde un usuario o rol especificados.
- No se puede conceder un permiso específico a grupos de usuarios.
- Al conceder o revocar permisos acotados, se modifican los permisos de todos los objetos actuales y futuros del ámbito.
- Los permisos limitados y lo permisos en el nivel de objeto operan de forma independiente. Por ejemplo, un usuario mantendrá los permisos en una tabla en los dos casos siguientes.
  - Al usuario se le concede el permiso SELECT en la tabla schema1.table1 y el permiso SELECT limitado en schema1. A continuación, el usuario revoca SELECT de todas las tablas del esquema schema1. El usuario retiene SELECT en schema1.table1.

  - Al usuario se le concede el permiso SELECT en la tabla schema1.table1 y el permiso SELECT limitado en schema1. A continuación, el usuario revoca SELECT de schema1.table1. El usuario retiene SELECT en schema1.table1.
- Para conceder o revocar permisos acotados, debe cumplir uno de los siguientes criterios:
  - Superusuarios
  - Usuarios con la opción de conceder ese permiso Para obtener más información sobre las opciones de concesión, vaya al parámetro WITH GRANT OPTION en GRANT.
- Los permisos acotados solo se pueden conceder o revocar a los objetos de la base de datos conectada o a las bases de datos importadas desde un recurso compartido de datos.
- Puede usar los permisos acotados para establecer los permisos predeterminados en una base de datos creada a partir de un recurso compartido de datos. Un usuario de un recurso compartido de datos del lado del consumidor al que se le conceden permisos acotados en una base de datos compartida los recibirá automáticamente para cualquier objeto nuevo que se añada al recurso compartido de datos del lado del productor.
- Los productores pueden conceder permisos específicos sobre los objetos dentro de un esquema a un recurso compartido de datos. (vista previa)

# Referencia de SQL

Con Amazon Redshift, puede aprovechar SQL para consultar y analizar de manera eficiente grandes cantidades de datos almacenados en el almacenamiento de datos. La referencia de SQL abarca la sintaxis y el uso de los comandos SQL, los tipos de datos, las funciones, los operadores, etc., lo que le permite extraer información y tomar decisiones basadas en los datos. Puede consultar esta referencia para escribir consultas optimizadas, crear y administrar objetos de bases de datos y realizar transformaciones de datos complejas. La siguiente referencia proporciona detalles completos sobre el uso de SQL en el entorno Amazon Redshift para cumplir los requisitos de análisis de datos.

Temas


- SQL en Amazon Redshift
- Uso de SQL
- Comandos SQL
- Referencia de funciones SQL
- Palabras reservadas


## SQL en Amazon Redshift

Temas


- Funciones SQL admitidas en el nodo principal
- Amazon Redshift y PostgreSQL


Amazon Redshift se desarrolla en torno a SQL estándar del sector, con la funcionalidad agregada de administrar conjuntos de datos muy grandes y admitir el análisis y la generación de informes de alto rendimiento de esos datos.

> **Note**
>
> El tamaño máximo de una sola instrucción SQL de Amazon Redshift es de 16 MB.

# Funciones SQL admitidas en el nodo principal

Algunas consultas de Amazon Redshift se distribuyen y ejecutan en los nodos de computación, y otras se ejecutan exclusivamente en el nodo principal.

El nodo principal distribuye SQL a los nodos de computación cuando una consulta utiliza como referencia tablas de sistemas o tablas creadas por el usuario (tablas con un prefijo STL o STV y vistas de sistema con un prefijo SVL o SVV). Una consulta que utiliza como referencia solamente tablas de catálogos (tablas con un prefijo PG, como por ejemplo PG_TABLE_DEF, que reside en el nodo principal) o que no utiliza ninguna tabla se ejecuta únicamente en el nodo principal.

Algunas funciones SQL de Amazon Redshift solo se admiten en el nodo principal, y no en los nodos informáticos. Una consulta que utiliza una función del nodo principal debe ejecutarse exclusivamente en ese nodo, no en los nodos informáticos, ya que se produciría un error.

En la documentación de cada función que debe ejecutarse exclusivamente en el nodo principal, se incluye una nota en la que se indica que la función producirá un error si utiliza como referencia tablas definidas por el usuario o tablas de sistema de Amazon Redshift. Consulte Funciones específicas del nodo principal para ver una lista de las funciones que se ejecutan exclusivamente en el nodo principal.

## Ejemplos

En los siguientes ejemplos se utiliza la base de datos de TICKIT de ejemplo. Para obtener más información acerca de las bases de datos de ejemplo, vaya a Base de datos de muestra.

CURRENT_SCHEMA

La función CURRENT_SCHEMA es una función específica de nodo principal. En este ejemplo, la consulta no hace referencia a una tabla, por lo que se ejecuta exclusivamente en el nodo principal.

```
select current_schema();

current_schema
---------------
public
```

En el siguiente ejemplo, la consulta hace referencia a una tabla de catálogo del sistema, por lo que se ejecuta exclusivamente en el nodo principal.

```
select * from pg_table_def
```

```
where schemaname = current_schema() limit 1;

 schemaname | tablename | column |   type   | encoding | distkey | sortkey | notnull
------------+-----------+--------+----------+----------+---------+---------+---------
 public     | category  | catid  | smallint | none     | t       |       1 | t
```

En el siguiente ejemplo, la consulta referencia una tabla de sistema de Amazon Redshift que se encuentra en los nodos informáticos, por lo que devuelve un error.

```
select current_schema(), userid from users;

INFO:  Function "current_schema()" not supported.
ERROR:  Specified types or functions (one per INFO message) not supported on Amazon
 Redshift tables.
```

SUBSTR

SUBSTR también es una función exclusiva para el nodo principal. En este ejemplo, la consulta no hace referencia a una tabla, por lo que se ejecuta exclusivamente en el nodo principal.

```
SELECT SUBSTR('amazon', 5);

+--------+
| substr |
+--------+
| on     |
+--------+
```

En el siguiente ejemplo, la consulta hace referencia a una tabla que se encuentra en los nodos de computación. Esto generará un error.

```
SELECT SUBSTR(catdesc, 1) FROM category LIMIT 1;

ERROR: SUBSTR() function is not supported (Hint: use SUBSTRING instead)
```

Para ejecutar correctamente la consulta anterior, utilice SUBSTRING.

```
SELECT SUBSTRING(catdesc, 1) FROM category LIMIT 1;

+--------------------------------+
|           substring            |
```

```
+---------------------------------+
| National Basketball Association |
+---------------------------------+
```

FACTORIAL()

FACTORIAL () es una función exclusiva para el nodo principal. En este ejemplo, la consulta no hace referencia a una tabla, por lo que se ejecuta exclusivamente en el nodo principal.

```
SELECT FACTORIAL(5);

 factorial
-------------
 120
```

En el siguiente ejemplo, la consulta hace referencia a una tabla que se encuentra en los nodos de computación. Esto provoca un error cuando se ejecuta mediante el editor de consultas v2.

```
create table t(a int);
insert into t values (5);
select factorial(a) from t;

ERROR: Specified types or functions (one per INFO message) not supported on Redshift
 tables.
Info: Function "factorial(bigint)" not supported.
```

# Amazon Redshift y PostgreSQL

Temas

- JDBC y ODBC de PostgreSQL y Amazon Redshift
- Características que se implementan de manera diferente
- Características no compatibles de PostgreSQL
- Tipos de datos de PostgreSQL no admitidos
- Funciones no compatibles de PostgreSQL

Amazon Redshift se basa en PostgreSQL. Amazon Redshift y PostgreSQL tienen una serie de diferencias muy importantes que debe tener en cuenta al diseñar y desarrollar aplicaciones de almacenamiento de datos.

Amazon Redshift está diseñado específicamente para aplicaciones de procesamiento analítico online (OLAP) e inteligencia empresarial (BI), que requieren consultas complejas en conjuntos de datos grandes. Dado que cumple requisitos muy diferentes, el esquema de almacenamiento de datos especializado y el motor de ejecución de consultas que Amazon Redshift utiliza son completamente diferentes a la implementación de PostgreSQL. Por ejemplo, mientras las aplicaciones de procesamiento de transacciones online (OLTP) normalmente almacenan los datos en filas, Amazon Redshift almacena los datos en columnas utilizando codificaciones de compresión de datos especializadas para lograr un uso óptimo de la memoria y las operaciones de E/S del disco. Se han omitido algunas características de PostgreSQL adecuadas para procesos de OLTP de menor escala para mejorar el rendimiento, como índices secundarios y operaciones eficientes de manipulación de los datos de una sola fila.

Consulte Arquitectura de Amazon Redshift para obtener una explicación más detallada de la arquitectura del sistema de almacenamiento de datos de Amazon Redshift.

PostgreSQL 9.x incluye algunas características que no son compatibles con Amazon Redshift. Además, hay diferencias importantes que debe conocer entre el SQL de Amazon Redshift y PostgreSQL. En esta sección, se ponen de manifiesto las diferencias entre Amazon Redshift y PostgreSQL, y se proporcionan directrices para desarrollar un almacenamiento de datos que aproveche al máximo la implementación del SQL de Amazon Redshift.

## JDBC y ODBC de PostgreSQL y Amazon Redshift

Como Amazon Redshift se basa en PostgreSQL, antes recomendábamos utilizar los controladores JDBC4 Postgresql en la versión 8.4.703 y los controladores psqlODBC en la versión 9.x. Si actualmente utiliza esos controladores, le recomendamos que opte por los nuevos controladores específicos de Amazon Redshift desde ahora. Para obtener más información acerca de los controladores y la configuración de las conexiones, consulte Controladores JDBC y ODBC para Amazon Redshift en la Guía de administración de Amazon Redshift.

Para evitar errores de falta de memoria del lado del cliente cuando recupera grandes conjuntos de datos a través de JDBC, puede habilitar el cliente para recuperar datos en lotes al configurar el parámetro de tamaño de búsqueda de JDBC. Para obtener más información, consulte Configuración del parámetro de tamaño de búsqueda de la JDBC.

Amazon Redshift no reconoce el parámetro maxRows de JDBC. En cambio, especifique una cláusula LIMIT para limitar el conjunto de resultados. También puede usar una cláusula OFFSET para ir a un determinado punto de partida en el conjunto de resultados.

## Características que se implementan de manera diferente

Muchos elementos del lenguaje SQL de Amazon Redshift tienen diferentes características de rendimiento y utilizan sintaxis y semántica distintas a las de la implementación de PostgreSQL equivalente.

> **Important**
>
> No suponga que la semántica de los elementos que Amazon Redshift y PostgreSQL tienen en común es idéntica. No olvide revisar la Guía para desarrolladores de Amazon Redshift para consultar Comandos SQL y así conocer las diferencias que a menudo son muy sutiles.

Un ejemplo en particular es el comando VACUUM, que se utiliza para limpiar y reorganizar tablas. VACUUM funciona de manera diferente y usa un conjunto diferente de parámetros que la versión PostgreSQL. Consulte Limpieza de tablas para obtener más información acerca del uso de VACUUM en Amazon Redshift.

A menudo, las características y herramientas de gestión y administración de bases de datos también son diferentes. Por ejemplo, Amazon Redshift mantiene un conjunto de vistas y tablas de sistema que proporcionan información sobre cómo funciona el sistema. Para obtener más información, consulta Vistas de monitoreo de SYS.

La siguiente lista incluye algunos ejemplos de características SQL que se implementan de manera diferente en Amazon Redshift.

- CREATE TABLE

  Amazon Redshift no admite espacios de tabla, particionamiento de tablas, herencia ni ciertas restricciones. La implementación de Amazon Redshift de CREATE TABLE le permite definir los algoritmos de ordenación y distribución para que las tablas optimicen el procesamiento paralelo.

  Amazon Redshift Spectrum admite el particionamiento de tablas a través del comando CREATE EXTERNAL TABLE.

- ALTER TABLE

  Solo se admite un subconjunto de acciones de ALTER COLUMN.

  ADD COLUMN admite la adición de una columna únicamente en cada instrucción ALTER TABLE.

- COPY

  El comando COPY de Amazon Redshift es muy especializado para habilitar la carga de datos desde buckets de Amazon S3 y tablas de Amazon DynamoDB, y para facilitar la compresión automática. Para obtener más información, consulte la sección Carga de datos en Amazon Redshift y la referencia del comando COPY.

- VACUUM

  Los parámetros para VACUUM son completamente diferentes. Por ejemplo, la operación VACUUM predeterminada en PostgreSQL simplemente recupera espacio y lo pone a disposición para volver a utilizarlo; sin embargo, la operación VACUUM predeterminada en Amazon Redshift es VACUUM FULL, que recupera espacio en el disco y reordena todas las filas.

- Se ignoran los espacios anteriores o posteriores en los valores VARCHAR cuando se comparan valores en cadena. Para obtener más información, consulte Importancia de los espacios en blancos anteriores y posteriores.

## Características no compatibles de PostgreSQL

Estas características de PostgreSQL no son compatibles con Amazon Redshift.

> **Important**
>
> No suponga que la semántica de los elementos que Amazon Redshift y PostgreSQL tienen en común es idéntica. No olvide revisar la Guía para desarrolladores de Amazon Redshift para consultar Comandos SQL y así conocer las diferencias que a menudo son muy sutiles.

- La herramienta de consulta psql no es compatible. Se admite el cliente de Amazon Redshift RSQL.
- Particiones de tabla (particiones de lista y rango)
- Espacios de tabla
- Restricciones
  - Único
  - Clave externa
  - Clave principal
  - Restricciones de comprobación

  - Restricciones de exclusión

    Se permiten la unicidad, las restricciones de claves principales y de claves externas, pero son solo para fines informativos. No están forzadas por el sistema, pero el planificador de consultas las usa.
- Herencia
- Columnas del sistema PostgresSQL

  El SQL de Amazon Redshift no define de manera implícita las columnas del sistema. No obstante, los siguientes nombres de columnas del sistema PostgreSQL no pueden utilizarse como nombres de columnas definidas por el usuario: `oid`, `tableoid`, `xmin`, `cmin`, `xmax`, `cmax` y `ctid`. Para obtener más información, consulte https://www.postgresql.org/docs/8.0/static/ddl-system-columns.html.
- Índices
- Cláusula NULLS en funciones de ventana
- Intercalaciones

  Amazon Redshift no admite secuencias de intercalación definidas por el usuario o específicas de una ubicación. Consulte Secuencias de intercalación.
- Expresiones de valor
  - Expresiones suscritas
  - Constructores de matriz
  - Constructores de fila
- Desencadenadores
- Administración de datos externos (SQL/MED)
- Funciones de tabla
- Lista VALUES utilizada como tablas de constantes
- Secuencias
- Búsqueda de texto completo
- Los permisos RULE y TRIGGER.

  Amazon Redshift concede o revoca estos permisos al ejecutar GRANT ALL o REVOKE ALL, pero la presencia o ausencia de los permisos RULE y TRIGGER no afecta en modo alguno los permisos de acceso del beneficiario.

## Tipos de datos de PostgreSQL no admitidos

Por lo general, si una consulta intenta utilizar un tipo de datos no admitido, incluidas formas explícitas o implícitas, devuelve un error. Sin embargo, algunas consultas que utilizan tipos de datos no admitidos se ejecutarán en el nodo principal, pero no en los nodos de computación. Consulte Funciones SQL admitidas en el nodo principal.

Para ver una lista de los tipos de datos admitidos, consulte Tipos de datos.

Estos tipos de datos de PostgreSQL no son compatibles con Amazon Redshift.

- Matrices
- BIT, BIT VARYING
- BYTEA
- Tipos compuestos
- Tipos enumerados
- Tipos geométricos (la implementación de tipos geométricos en Amazon Redshift es diferente a la de PostgreSQL)
- HSTORE
- JSON
- Tipos de direcciones de red
- Tipos numéricos
  - SERIAL, BIGSERIAL, SMALLSERIAL
  - MONEY
- Tipos de Identificadores de objeto
- Pseudo-tipos
- Tipos de rangos
- Tipos de carácter especial
  - "char": un tipo interno de un solo byte (donde el tipo de datos llamado char se escribe entre comillas).
  - name: un tipo interno para nombres de objeto.

  Para más información acerca de estos tipos, vea Tipos de caracter especial en la documentación de PostgreSQL.

- Tipos de búsquedas de texto
- TXID_SNAPSHOT
- UUID
- XML

## Funciones no compatibles de PostgreSQL

Muchas funciones que no se excluyen tienen diferentes semánticas o usos. Por ejemplo, algunas funciones admitidas se ejecutarán solo en el nodo principal. También, algunas funciones no admitidas devolverán un error cuando se ejecutan en el nodo principal. El hecho de que estas funciones no devuelven un error en algunos casos no debe considerarse como una indicación de que la función es compatible con Amazon Redshift.

> **Important**
>
> No suponga que la semántica de los elementos que Amazon Redshift y PostgreSQL tienen en común es idéntica. No olvide revisar la Guía para desarrolladores de bases de datos de Amazon Redshift para consultar Comandos SQL y así conocer las diferencias que a menudo son muy sutiles.

Para obtener más información, consulte Funciones SQL admitidas en el nodo principal.

Estas funciones de PostgreSQL no son compatibles con Amazon Redshift.

- Funciones de investigación de privilegios de acceso
- Funciones de bloqueo consultivo
- Funciones de agregación
  - STRING_AGG()
  - ARRAY_AGG()
  - EVERY()
  - XML_AGG()
  - CORR()
  - COVAR_POP()
  - COVAR_SAMP()

  - REGR_AVGX(), REGR_AVGY()
  - REGR_COUNT()
  - REGR_INTERCEPT()
  - REGR_R2()
  - REGR_SLOPE()
  - REGR_SXX(), REGR_SXY(), REGR_SYY()
- Funciones y operadores de matriz
- Funciones de control de backup
- Funciones de información de comentarios
- Funciones de ubicación de objetos de la base de datos
- Funciones de tamaño de objetos de la base de datos
- Operadores y funciones de fecha y hora
  - CLOCK_TIMESTAMP()
  - JUSTIFY_DAYS(), JUSTIFY_HOURS(), JUSTIFY_INTERVAL()
  - PG_SLEEP()
  - TRANSACTION_TIMESTAMP()
- Funciones de soporte de ENUM
- Operadores y funciones geométricas
- Funciones de acceso de archivo genérico
- IS DISTINCT FROM
- Operadores y funciones de direcciones de red
- Funciones matemáticas
  - DIV()
  - SETSEED()
  - WIDTH_BUCKET()
- Funciones de devolución de conjuntos
  - GENERATE_SERIES()
  - GENERATE_SUBSCRIPTS()
- Operadores y funciones de rango
- Funciones de control de recuperación

- Funciones de información de recuperación
- Función ROLLBACK TO SAVEPOINT
- Funciones de investigación de visibilidad de esquemas
- Funciones de señalización de servidores
- Funciones de sincronización de instantáneas
- Funciones de manipulación de secuencias
- Funciones de cadena
  - BIT_LENGTH()
  - OVERLAY()
  - CONVERT(), CONVERT_FROM(), CONVERT_TO()
  - ENCODE()
  - FORMAT()
  - QUOTE_NULLABLE()
  - REGEXP_MATCHES()
  - REGEXP_SPLIT_TO_ARRAY()
  - REGEXP_SPLIT_TO_TABLE()
- Funciones de información de catálogo del sistema
- Funciones de información del sistema
  - CURRENT_CATALOG CURRENT_QUERY()
  - INET_CLIENT_ADDR()
  - INET_CLIENT_PORT()
  - INET_SERVER_ADDR() INET_SERVER_PORT()
  - PG_CONF_LOAD_TIME()
  - PG_IS_OTHER_TEMP_SCHEMA()
  - PG_LISTENING_CHANNELS()
  - PG_MY_TEMP_SCHEMA()
  - PG_POSTMASTER_START_TIME()
  - PG_TRIGGER_DEPTH()
  - SHOW VERSION()



- Operadores y funciones de búsqueda de texto

- Funciones de instantáneas e ID de transacciones
- Funciones de disparador
- Funciones XML

# Uso de SQL

Temas


- Convenciones de la referencia de SQL
- Elementos básicos
- Expressions
- Condiciones


El lenguaje SQL consta de comandos y funciones que se utilizan para trabajar con bases de datos y objetos de base de datos. El lenguaje también aplica reglas en relación al uso de tipos de datos, expresiones y literales.

## Convenciones de la referencia de SQL

En esta sección se explican las convenciones que se utilizan para escribir la sintaxis para las expresiones, los comandos y las funciones SQL que se describen en la sección de referencia de SQL.

| Carácter | Descripción |
|---|---|
| CAPS | Las palabras en mayúscula son palabras clave. |
| [ ] | Los corchetes denotan argumentos opcionales. Varios argumentos entre corchetes indican que puede seleccionar cualquier cantidad de argumento s. Además, los argumentos entre corchetes en líneas distintas indican que el analizador de Amazon Redshift espera que los argumentos estén en el orden en que aparecen en la sintaxis. Para ver un ejemplo, consulta SELECT. |
| { } | Las llaves indican que debe seleccionar uno de los argumentos contenidos en las llaves. |
| \| | Las barras verticales indican que puede seleccionar entre los argumentos. |

| Carácter | Descripción |
| --- | --- |
| cursiva | Las palabras en cursiva indican marcadores de posición. Debe insertar el valor adecuado en lugar de la palabra en cursiva. |
| . . . | Los puntos suspensivos indican que puede repetir el elemento anterior. |
| ' | Las palabras entre comillas simples indican que debe escribir las comillas. |

# Elementos básicos

Temas

- Nombres e identificadores
- Literales
- Nulos
- Tipos de datos
- Secuencias de intercalación

En esta sección se cubren las reglas para trabajar con nombres de objetos de base de datos, literales, nulos y tipos de datos.

## Nombres e identificadores

Los nombres identifican objetos de base de datos, incluidas tablas y columnas, así como usuarios y contraseñas. Los términos nombre e identificador se pueden usar indistintamente. Hay dos tipos de identificadores: identificadores estándar e identificadores delimitados o entre comillas. Los identificadores solo deben tener caracteres UTF-8 imprimibles. Las letras ASCII en los identificadores estándar y delimitados distinguen entre mayúsculas y minúsculas, y se incorporan en minúsculas en la base de datos. En los resultados de la consulta, los nombres de columna se devuelven en minúsculas de forma predeterminada. Para devolver nombres de columna en mayúsculas, establezca el parámetro de configuración describe_field_name_in_uppercase en **`true`**.

Identificadores estándar

Los identificadores SQL estándar se adhieren a un conjunto de reglas y:

- Deben comenzar con un carácter alfabético de byte único ASCII o carácter de guion bajo, o un carácter multibyte UTF-8 de dos a cuatro bytes.
- Los caracteres posteriores pueden ser caracteres alfanuméricos de byte único ASCII, guiones bajos o signos de dólar, o caracteres multibyte UTF-8 de dos a cuatro bytes.
- Deben tener entre 1 y 127 bytes, sin incluir las comillas para los identificadores delimitados.
- No deben contener comillas ni espacios.
- No deben ser una palabra clave SQL reservada.

### Identificadores delimitados

Los identificadores delimitados (también conocidos como identificadores entre comillas) comienzan y finalizan con comillas dobles ("). Si utiliza un identificador delimitado, debe usar las comillas dobles para cada referencia a ese objeto. El identificador puede contener cualquier carácter UTF-8 imprimible estándar, aparte de las comillas dobles. Por lo tanto, puede crear nombres de columna o tabla que incluyan caracteres de otra forma ilegales, como espacios y el símbolo de porcentaje.

Las letras ASCII en los identificadores delimitados distinguen entre mayúsculas y minúsculas, y se incorporan en minúsculas. Para utilizar una comilla doble en una cadena, debe precederla con otro carácter de comilla doble.

### Identificadores que distinguen entre mayúsculas y minúsculas

Los identificadores que distinguen entre mayúsculas y minúsculas (también conocidos como identificadores con combinación de mayúsculas y minúsculas) pueden contener letras tanto mayúsculas como minúsculas. Para utilizar identificadores que distinguen entre mayúsculas y minúsculas, puede establecer la configuración `enable_case_sensitive_identifier` en `true`. Puede establecer esta configuración para el clúster o para una sesión. Para obtener más información, consulte Valores de parámetros predeterminados en la Guía de administración de Amazon Redshift y enable_case_sensitive_identifier.

### Nombres de columna del sistema

Los siguientes nombres de columnas del sistema de PostgreSQL no pueden utilizarse como nombres de columnas en columnas definidas por el usuario. Para obtener más información, consulte https://www.postgresql.org/docs/8.0/static/ddl-system-columns.html.

- `oid`

- tableoid
- xmin
- cmin
- xmax
- cmax
- ctid

Ejemplos

En esta tabla se muestran ejemplos de identificadores delimitados, el resultado y una explicación:

| Sintaxis | Resultado | Explicación |
|---|---|---|
| "group" | grupo | GROUP es una palabra reservada, por lo que su uso dentro de un identificador requiere comillas dobles. |
| """WHERE""" | "where" | WHERE también es una palabra reservada. Para incluir comillas en la cadena, aplique un escape a cada carácter de comilla doble con caracteres de comillas dobles adicionales. |
| "This name" | this name | Se requieren comillas dobles para preservar el espacio. |
| "This ""IS IT""" | this "is it" | Las comillas que encierran IS IT deben estar precedidas por una comilla adicional para que pasen a formar parte del nombre. |

Para crear una tabla denominada grupo con una columna denominada "is it":

```
create table "group" (
"This ""IS IT""" char(10));
```

Las siguientes consultas devuelven el mismo resultado:

```
select "This ""IS IT"""
from "group";
```

```
this "is it"
--------------
(0 rows)
```

```
select "this ""is it"""
from "group";

this "is it"
--------------
(0 rows)
```

La siguiente sintaxis `table.column` completamente calificada también devuelve el mismo resultado:

```
select "group"."this ""is it"""
from "group";

this "is it"
--------------
(0 rows)
```

El siguiente comando CREATE TABLE crea una tabla con una barra diagonal en un nombre de columna:

```
create table if not exists city_slash_id(
                  "city/id" integer not null,
                  state char(2) not null);
```

## Literales

Un literal o una constante es un valor de dato fijo que está compuesto por una secuencia de caracteres o una constante numérica. Amazon Redshift admite varios tipos de literales, entre los que se incluyen los siguientes:

- Literales numéricos para enteros, decimales y números en coma flotante. Para obtener más información, consulte Literales enteros y con coma flotante.
- Literales de caracteres, también conocidos como cadenas, cadenas de caracteres y constantes de caracteres

- Literales de intervalo y de fecha y hora, utilizados como tipos datos de fecha y hora. Para obtener más información, consulte Literales de fecha, hora y marca temporal y Literales y tipos de datos de intervalo.

## Nulos

Si hay una columna en una fila que falta, es desconocida o no aplica, es un valor nulo o se dice que contiene un valor nulo. Los nulos pueden aparecer en campos de cualquier tipo de datos que no están limitados por restricciones NOT NULL o clave principal. Un nulo no equivale al valor cero o a una cadena vacía.

Cualquier expresión aritmética que contiene un nulo siempre toma el valor de un nulo. Todos los operadores, excepto la concatenación, devuelven un nulo cuando se proporciona un operando o argumento nulo.

Para probar nulos, utilice las condiciones de comparación IS NULL e IS NOT NULL. Dado que el valor nulo representa una falta de datos, un nulo no es igual o desigual a cualquier valor o a otro nulo.

## Tipos de datos

Temas

- Caracteres multibyte
- Tipos numéricos
- Tipos de caracteres
- Tipos de fecha y hora
- Tipo booleano
- Tipo HLLSKETCH
- Tipo SUPER
- Tipo VARBYTE
- Conversión y compatibilidad de tipos

Cada valor que Amazon Redshift almacena o recupera tiene un tipo de datos con un conjunto fijo de propiedades asociadas. Los tipos de datos se declaran cuando se crean las tablas. Un tipo de datos limita el conjunto de valores que una columna o un argumento puede contener.

En la siguiente tabla, se enumeran los tipos de datos que puede usar en tablas de Amazon Redshift.

| Tipo de datos: | Alias | Descripción |
|---|---|---|
| SMALLINT | INT2 | Entero firmado de dos bytes |
| INTEGER | INT, INT4 | Entero firmado de cuatro bytes |
| BIGINT | INT8 | Entero firmado de ocho bytes |
| DECIMAL | NUMERIC | Numérico exacto de precisión seleccionable |
| REAL | FLOAT4 | Número en coma flotante de precisión única |
| DOUBLE PRECISION | FLOAT8, FLOAT | Número en coma flotante de precisión doble |
| CHAR | CHARACTER, NCHAR, BPCHAR | Cadena de caracteres de longitud fija |
| VARCHAR | CHARACTER VARYING, NVARCHAR, TEXT | Cadena de caracteres de longitud variable con un límite definido por el usuario |
| DATE | | Fecha de calendario (año, mes, día) |
| TIME | TIME WITHOUT TIME ZONE | Hora del día |
| TIMETZ | HORA CON ZONA HORARIA | Hora del día con zona horaria |
| TIMESTAMP | TIMESTAMP WITHOUT TIME ZONE | Fecha y hora (sin zona horaria) |
| TIMESTAMPTZ | MARCA DE TIEMPO CON ZONA HORARIA | Fecha y hora (con zona horaria) |

| Tipo de datos: | Alias | Descripción |
|---|---|---|
| INTERVALO AÑO A MES | | Duración del tiempo en orden de año a mes |
| INTERVALO DÍA A SEGUNDO | | Duración del tiempo en orden de día a segundo |
| BOOLEAN | BOOL | Booleano lógico (true/false) |
| HLLSKETCH | | Tipo utilizado con bocetos de HyperLogLog |
| SUPER | | Un tipo de datos superconj unto que engloba todos los tipos escalares de Amazon Redshift, incluidos aquellos más complejos como ARRAY y STRUCTS. |
| VARBYTE | VARBINARY, BINARY VARYING | Valor binario de longitud variable |
| GEOMETRY | | Datos espaciales |
| GEOGRAPHY | | Datos espaciales |

> **Note**
>
> Para más información acerca de tipos de datos no compatibles, como "char" (vea que char está entre comillas), vea Tipos de datos de PostgreSQL no admitidos.

### Caracteres multibyte

El tipo de datos VARCHAR es compatible con caracteres multibyte UTF-8 de hasta un máximo de cuatro bytes. Los caracteres de cinco bytes o más no son compatibles. Para calcular el tamaño de una columna VARCHAR que contiene caracteres multibyte, multiplique el número de caracteres por

el número de bytes por carácter. Por ejemplo, si una cadena tiene cuatro caracteres chinos y cada carácter tiene tres bytes, necesitará una columna VARCHAR(12) para almacenar la cadena.

El tipo de datos VARCHAR no es compatible con los siguientes valores de punto UTF-8 no válidos:

`0xD800 – 0xDFFF` (Secuencias de bytes: `ED A0 80` a `ED BF BF`)

El tipo de datos CHAR no es compatible con los caracteres multibyte.

Tipos numéricos

Temas

- Tipos de enteros
- Tipo DECIMAL o NUMERIC
- Notas acerca del uso de las columnas DECIMAL o NUMERIC de 128 bits
- Tipos de números en coma flotante
- Cómputos con valores numéricos
- Literales enteros y con coma flotante
- Ejemplos con tipos numéricos

Los tipos de datos numéricos incluyen enteros, decimales y números en coma flotante.

Tipos de enteros

Use los tipos de datos SMALLINT, INTEGER y BIGINT para almacenar los números completos de varios rangos. No puede almacenar valores fuera del rango permitido para cada tipo.

| Nombre | Almacenamiento | Range |
|---|---|---|
| SMALLINT o INT2 | 2 bytes | De -32768 a +32767 |
| INTEGER, INT o INT4 | 4 bytes | De -2147483648 a 2147483647 |
| BIGINT o INT8 | 8 bytes | De -92233720 36854775808 a 922337203 6854775807 |

### Tipo DECIMAL o NUMERIC

Use el tipo de datos DECIMAL o NUMERIC para almacenar valores con una precisión definida por el usuario. Las palabras clave DECIMAL y NUMERIC son intercambiables. En este documento, decimal es el término preferido para este tipo de datos. El término numérico se utiliza genéricamente para referirse a tipos de datos enteros, decimales y con coma flotante.

| Almacenamiento | Range |
|---|---|
| Variable, hasta 128 bits para tipos DECIMAL sin comprimir. | Los enteros firmados de 128 bits con hasta 38 dígitos de precisión. |

Defina una columna DECIMAL en una tabla especificando una precisión y una escala:

```
decimal(precision, scale)
```

precisión

El número total de dígitos significativos en todo el valor: la cantidad de dígitos de ambos lados del punto decimal. Por ejemplo, el número `48.2891` tiene una precisión de 6 y una escala de 4. La precisión predeterminada es 18, si no se especifica. La precisión máxima es 38.

Si la cantidad de dígitos a la izquierda del punto decimal en un valor de entrada supera la precisión de la columna menos su escala, no se puede copiar el valor a la columna (ni insertar ni actualizar). Esta regla se aplica a cualquier valor que caiga fuera del rango de la definición de la columna. Por ejemplo, el rango permitido de valores para una columna `numeric(5,2)` es de `-999.99` a `999.99`.

scale

El número de dígitos decimales en la parte fraccional del valor, a la derecha del punto decimal. Los enteros tienen una escala de cero. En la especificación de una columna, el valor de la escala debe ser inferior que o igual al valor de precisión. La escala predeterminada es 0, si no se especifica. La escala máxima es 37.

Si la escala de un valor de entrada que se carga en una tabla es mayor que la escala de la columna, el valor se redondea a la escala especificada. Por ejemplo, la columna PRICEPAID de la tabla SALES es una columna DECIMAL(8,2). Si se inserta un valor DECIMAL(8,4) en la columna PRICEPAID, el valor se redondea a una escala de 2.

```
insert into sales
values (0, 8, 1, 1, 2000, 14, 5, 4323.8951, 11.00, null);

select pricepaid, salesid from sales where salesid=0;

pricepaid | salesid
-----------+---------
4323.90 |       0
(1 row)
```

Sin embargo, no se redondean los resultados de formas explícitas de los valores seleccionados de tablas.

> **Note**
>
> El valor positivo máximo que puede insertar en una columna DECIMAL(19,0) es `9223372036854775807` ($2^{63}$ -1). El valor negativo máximo es `-9223372036854775808`. Por ejemplo, un intento de insertar el valor `9999999999999999` (19 nueves) provocará un error de desbordamiento. Independientemente de la ubicación de la coma decimal, la cadena de mayor tamaño que Amazon Redshift puede representar como un número DECIMAL es `9223372036854775807`. Por ejemplo, el valor más grande que puede cargar en una columna DECIMAL(19,18) es `9.223372036854775807`.
> Estas reglas se deben a que los valores DECIMAL con 19 o menos dígitos significativos de precisión se almacenan internamente como números enteros de 8 bytes, mientras que los valores DECIMAL con una precisión de 20 a 38 dígitos significativos se almacenan como enteros de 16 bytes.

Notas acerca del uso de las columnas DECIMAL o NUMERIC de 128 bits

No asigne arbitrariamente la precisión máxima de las columnas DECIMAL a menos que esté seguro de que la aplicación requiere esa precisión. Los valores de 128 bits utilizan el doble de espacio en el disco en comparación de los valores de 64 bits y pueden alargar el tiempo de ejecución de la consulta.

Tipos de números en coma flotante

Use el tipo de datos REAL o DOUBLE PRECISION para almacenar valores numéricos con precisión variable. Estos tipos son inexactos, lo que significa que algunos valores se almacenan como

aproximaciones, por lo que puede haber pequeñas discrepancias al almacenar y devolver un valor específico. Si requiere almacenamiento y cálculos exactos (como para importes monetarios), use el tipo de datos DECIMAL.

REAL representa el formato de coma flotante de precisión simple, según la norma IEEE 754 para la aritmética binaria de coma flotante. Tiene una precisión de unos 6 dígitos y un intervalo de 1E-37 a 1E+37 aproximadamente. También puede especificar este tipo de datos como FLOAT4.

DOUBLE PRECISION representa el formato de coma flotante de doble precisión, según la norma IEEE 754 para la aritmética binaria de coma flotante. Tiene una precisión de unos 15 dígitos y un intervalo de 1E-307 a 1E+308 aproximadamente. También puede especificar este tipo de datos como FLOAT o FLOAT8.

Además de los valores numéricos normales, los tipos de punto flotante tienen varios valores especiales. Utilice estos valores entre comillas simples cuando los utilice en SQL:

- `NaN`: no es un número
- `Infinity`: infinito
- `-Infinity`: infinito negativo

Por ejemplo, para insertar no es un número en la columna `day_charge` de la tabla `customer_activity`, ejecute el siguiente código SQL:

```
insert into customer_activity(day_charge) values('NaN');
```

Cómputos con valores numéricos

En este contexto, cómputo se refiere a operaciones matemáticas binarias: suma, resta, multiplicación y división. En esta sección se describen los tipos devueltos previstos para estas operaciones, así como la fórmula específica que se aplica para determinar la precisión y la escala cuando hay tipos de datos DECIMAL involucrados.

Cuando se computan los valores numéricos durante el procesamiento de consultas, puede encontrar casos donde el cómputo no es posible y la consulta devuelve un error de desbordamiento numérico. También puede encontrar casos donde una escala de valores computados varía o es inesperada. Para algunas operaciones, puede usar conversiones explícitas (tipo de promoción) o parámetros de configuración de Amazon Redshift para solucionar estos problemas.

Para obtener más información acerca de los resultados de cálculo similares con funciones SQL, consulte Funciones de agregación.

Tipos devueltos para cómputos

Dado el conjunto de tipos de datos numéricos admitidos en Amazon Redshift, en la siguiente tabla, se muestran los tipos devueltos previstos para las operaciones de suma, resta, multiplicación y división. La primera columna del lado izquierdo de la tabla representa el primer operando del cálculo, y la fila superior representa el segundo operando.

| Operando 1 | Operando 2 | Tipo de retorno |
|---|---|---|
| INT2 | INT2 | INT2 |
| INT2 | INT4 | INT4 |
| INT2 | INT8 | INT8 |
| INT2 | DECIMAL | DECIMAL |
| INT2 | FLOAT4 | FLOAT8 |
| INT2 | FLOAT8 | FLOAT8 |
| INT4 | INT4 | INT4 |
| INT4 | INT8 | INT8 |
| INT4 | DECIMAL | DECIMAL |
| INT4 | FLOAT4 | FLOAT8 |
| INT4 | FLOAT8 | FLOAT8 |
| INT8 | INT8 | INT8 |
| INT8 | DECIMAL | DECIMAL |
| INT8 | FLOAT4 | FLOAT8 |
| INT8 | FLOAT8 | FLOAT8 |

| Operando 1 | Operando 2 | Tipo de retorno |
|---|---|---|
| DECIMAL | DECIMAL | DECIMAL |
| DECIMAL | FLOAT4 | FLOAT8 |
| DECIMAL | FLOAT8 | FLOAT8 |
| FLOAT4 | FLOAT8 | FLOAT8 |
| FLOAT8 | FLOAT8 | FLOAT8 |

Precisión y escala de resultados DECIMAL computados

En la siguiente tabla se resumen las reglas para computar la precisión y la escala resultantes cuando las operaciones matemáticas devuelven resultados DECIMAL. En esta tabla, `p1` y `s1` representan la precisión y escala del primer operando en un cálculo, y `p2` y `s2` representan la precisión y escala del segundo operando. (Independientemente de estos cálculos, la precisión de resultados máxima es 38 y la escala de resultados máxima es 38).

| Operación | Precisión y escala del resultado |
|---|---|
| + o bien - | Escalado = `max(s1,s2)`<br>Precisión = `max(p1-s1,p2-s2)+1+scale` |
| * | Escalado = `s1+s2`<br>Precisión = `p1+p2+1` |
| / | Escalado = `max(4,s1+p2-s2+1)`<br>Precisión = `p1-s1+ s2+scale` |

Por ejemplo, las columnas PRICEPAID y COMMISSION de la tabla SALES son columnas DECIMAL(8,2). Si divide PRICEPAID por COMMISSION (o viceversa), la fórmula se aplica de la siguiente manera:

```
Precision = 8-2 + 2 + max(4,2+8-2+1)
```

```
= 6 + 2 + 9 = 17

Scale = max(4,2+8-2+1) = 9

Result = DECIMAL(17,9)
```

El siguiente cálculo es la regla general para computar la precisión y la escala resultantes para operaciones realizadas en valores DECIMAL con operadores como UNION, INTERSECT o EXCEPT, o funciones como COALESCE y DECODE:

```
Scale = max(s1,s2)
Precision = min(max(p1-s1,p2-s2)+scale,19)
```

Por ejemplo, una tabla DEC1 con una columna DECIMAL(7,2) se combina con una tabla DEC2 con una columna DECIMAL(15,3) para crear una tabla DEC3. En el esquema de DEC3 se muestra que la columna se convierte en una columna NUMERIC(15,3).

```
create table dec3 as select * from dec1 union select * from dec2;
```

Resultado

```
select "column", type, encoding, distkey, sortkey
from pg_table_def where tablename = 'dec3';

column |     type      | encoding | distkey | sortkey
-------+---------------+----------+---------+---------
c1     | numeric(15,3) | none     | f       | 0
```

En el ejemplo anterior, la fórmula se aplica de la siguiente manera:

```
Precision = min(max(7-2,15-3) + max(2,3), 19)
= 12 + 3 = 15

Scale = max(2,3) = 3

Result = DECIMAL(15,3)
```

Notas sobre las operaciones de división

Para las operaciones de división, dividir las condiciones por cero devuelve errores.

El límite de escala de 100 se aplica después de que se calculan la precisión y la escala. Si la escala resultante calculada es superior a 100, los resultados de la división están escalados de la siguiente manera:

- Precisión = `precision - (scale - max_scale)`
- Escalado = `max_scale`

Si la precisión calculada es superior a la precisión máxima (38), la precisión se reduce a 38 y la escala se convierte en el resultado de: `max((38 + scale - precision), min(4, 100))`

Condiciones de desbordamiento

Se revisa el desbordamiento para todos los cómputos numéricos. Los datos DECIMAL con una precisión de 19 o menos se almacenan como enteros de 64 bits. Los datos DECIMAL con una precisión superior a 19 se almacenan como enteros de 128 bits. La precisión máxima para todos los valores DECIMAL es 38 y la escala máxima es 37. Los errores de desbordamiento ocurren cuando un valor supera estos límites, que se aplican en los conjuntos de resultados intermedios y finales:

- Las formas explícitas generan errores de desbordamiento de tiempo de ejecución cuando los valores de datos específicos no se ajustan a la precisión o escala solicitada y especificada por la función de formación. Por ejemplo, no puede formar todos los valores de la columna PRICEPAID en la tabla SALES (una columna DECIMAL(38,2)) y devolver un resultado DECIMAL(7,3):

  ```
  select pricepaid::decimal(7,3) from sales;
  ERROR:  Numeric data overflow (result precision)
  ```

  Este error se produce porque algunos de los valores más grandes de la columna PRICEPAID no se pueden formar.
- Las operaciones de multiplicación producen resultados en los que la escala de resultados es la suma de la escala de cada operando. Si ambos operandos tienen una escala de 4, por ejemplo, la escala resultante es 8, dejando solo 10 dígitos para el lado izquierdo del punto decimal. Por lo tanto, es relativamente fácil encontrarse con condiciones de desbordamiento cuando multiplica dos números grandes que tienen escalas significativas.

  El siguiente ejemplo produce un error de desbordamiento.

  ```
  SELECT CAST(1 AS DECIMAL(38, 20)) * CAST(10 AS DECIMAL(38, 20));
  ERROR: 128 bit numeric data overflow (multiplication)
  ```

Puede evitar el error de desbordamiento mediante la división en lugar de la multiplicación. Utilice el siguiente ejemplo para dividir por 1 dividido por el divisor original.

```
SELECT CAST(1 AS DECIMAL(38, 20)) / (1 / CAST(10 AS DECIMAL(38, 20)));
+----------+
| ?column? |
+----------+
| 10       |
+----------+
```

Cálculos numéricos con tipos INTEGER y DECIMAL

Cuando uno de los operandos en un cálculo tiene un tipo de datos INTEGER y el otro operando es DECIMAL, el operando INTEGER se forma implícitamente como un DECIMAL:

- INT2 (SMALLINT) se forma como DECIMAL(5,0)
- INT4 (INTEGER) se forma como DECIMAL(10,0)
- INT8 (BIGINT) se forma como DECIMAL(19,0)

Por ejemplo, si multiplica SALES.COMMISSION, una columna DECIMAL(8,2), y SALES.QTYSOLD, una columna SMALLINT, este cálculo se forma de la siguiente manera:

```
DECIMAL(8,2) * DECIMAL(5,0)
```

Literales enteros y con coma flotante

Los literales o las constantes que representan números pueden ser enteros o números en coma flotante.

Literales enteros

La constante de un número entero es una secuencia de los dígitos del 0 al 9, con un signo opcional positivo (+) o negativo (-) antes de los dígitos.

Sintaxis

```
[ + | - ] digit ...
```

### Ejemplos

Entre los enteros válidos se incluyen:

```
23
-555
+17
```

### Literales en coma flotante

Los literales en coma flotante (también conocidos como literales fraccionales, numéricos o decimales) son secuencias de dígitos que incluyen un punto decimal y, opcionalmente, el marcador de exponente (e).

### Sintaxis

```
[ + | - ] digit ... [ . ] [ digit ...]
[ e | E [ + | - ] digit ... ]
```

### Argumentos

e | E

e o E indica que el número está especificado en notación científica.

### Ejemplos

Entre los literales en coma flotante válidos se incluyen:

```
3.14159
-37.
2.0e19
-2E-19
```

### Ejemplos con tipos numéricos

### Instrucción CREATE TABLE

La siguiente instrucción CREATE TABLE demuestra la declaración de diferentes tipos de datos numéricos:

```
create table film (
film_id integer,
language_id smallint,
original_language_id smallint,
rental_duration smallint default 3,
rental_rate numeric(4,2) default 4.99,
length smallint,
replacement_cost real default 25.00);
```

Intento de insertar un entero fuera de rango

En el siguiente ejemplo se intenta insertar el valor 33000 en una columna SMALLINT.

```
insert into film(language_id) values(33000);
```

El rango de SMALLINT es de -32768 a +32767, por lo que Amazon Redshift devuelve un error.

```
An error occurred when executing the SQL command:
insert into film(language_id) values(33000)

ERROR: smallint out of range [SQL State=22003]
```

Inserción de un valor decimal en una columna de valores enteros

En el siguiente ejemplo se inserta un valor decimal en una columna INT.

```
insert into film(language_id) values(1.5);
```

Este valor se inserta pero se redondea al valor entero de 2.

Inserción correcta de un decimal debido a que se redondea su escala

En el siguiente ejemplo se inserta un valor decimal que tiene una precisión más grande que la columna.

```
insert into film(rental_rate) values(35.512);
```

En este caso, se inserta el valor `35.51` en la columna.

Intento de insertar un valor decimal fuera de rango

En este caso, el valor `350.10` está fuera de rango. La cantidad de dígitos para los valores de las columnas DECIMAL es igual a la precisión de la columna menos su escala (4 menos 2 para la columna RENTAL_RATE). En otras palabras, el rango permitido para una columna `DECIMAL(4,2)` es de `-99.99` a `99.99`.

```
insert into film(rental_rate) values (350.10);
ERROR:  numeric field overflow
DETAIL:  The absolute value is greater than or equal to 10^2 for field with precision
 4, scale 2.
```

Inserción de valores de precisión variable en una columna REAL

En el siguiente ejemplo se insertan valores de precisión variable en una columna REAL.

```
insert into film(replacement_cost) values(1999999.99);

insert into film(replacement_cost) values(1999.99);

select replacement_cost from film;

+------------------+
| replacement_cost |
+------------------+
| 2000000          |
| 1999.99          |
+------------------+
```

El valor `1999999.99` se convierte a `2000000` para cumplir con el requisito de precisión para la columna `REAL`. El valor `1999.99` se carga como está.

Tipos de caracteres

Temas

- Almacenamiento y rangos
- CHAR o CHARACTER
- VARCHAR o CHARACTER VARYING
- Tipos NCHAR y NVARCHAR
- Tipos TEXT y BPCHAR

- Importancia de los espacios en blancos anteriores y posteriores
- Ejemplos con tipos de caracteres


Los tipos de datos de caracteres incluyen CHAR (carácter) y VARCHAR (carácter variable).

Almacenamiento y rangos

Los tipos de datos CHAR y VARCHAR se definen en términos de bytes, no de caracteres. Una columna CHAR solo puede contener caracteres de un byte, por lo que una columna CHAR(10) puede contener una cadena con una longitud máxima de 10 bytes. Un VARCHAR puede contener caracteres multibyte de hasta un máximo de cuatro bytes por carácter. Por ejemplo, una columna VARCHAR(12) puede contener 12 caracteres de un byte, 6 caracteres de dos bytes, 4 caracteres de tres bytes o 3 caracteres de cuatro bytes.

| Nombre | Almacenamiento | Rango (ancho de columna) |
|---|---|---|
| CHAR, CHARACTER o NCHAR | Longitud de la cadena, incluidos espacios en blanco anteriores o posteriores (si corresponde) | 4 096 bytes |
| VARCHAR, CHARACTER VARYING o NVARCHAR | 4 bytes + bytes totales por caracteres, donde cada carácter puede tener entre 1 y 4 bytes. | 65 535 bytes (64K -1) |
| BPCHAR | Convertido a CHAR(256) de longitud fija. | 256 bytes |
| TEXT | Convertido a VARCHAR(256). | 260 bytes |

> **Note**
>
> La sintaxis de CREATE TABLE admite la palabra clave MAX para los tipos de datos de caracteres. Por ejemplo:
>
> ```
> create table test(col1 varchar(max));
> ```
>
> El ajuste MAX define el ancho de la columna como 4 096 bytes para CHAR o 65 535 bytes para VARCHAR.

CHAR o CHARACTER

Utilice una columna CHAR o CHARACTER para almacenar cadenas de longitud fija. Estas cadenas está rellenadas con espacios en blanco, por lo que una columna CHAR(10) siempre ocupa 10 bytes de almacenamiento.

```
char(10)
```

Una columna CHAR sin una especificación de longitud resulta en una columna CHAR(1).

VARCHAR o CHARACTER VARYING

Utilice una columna VARCHAR o VARYING CHARACTER para almacenar cadenas de longitud variable con un límite fijo. Estas cadenas no se rellenan con espacios en blancos, por lo que una columna VARCHAR(120) consta de un máximo de 120 caracteres de un byte, 60 caracteres de dos bytes, 40 caracteres de tres bytes o 30 caracteres de cuatro bytes.

```
varchar(120)
```

Si utiliza el tipo de datos VARCHAR sin un especificador de longitud en una instrucción CREATE TABLE, la longitud predeterminada es 256. Si se utiliza en una expresión, el tamaño de la salida se determina mediante la expresión de entrada (hasta 65535).

Tipos NCHAR y NVARCHAR

Puede crear columnas con los tipos NCHAR y NVARCHAR (también conocidos como los tipos NATIONAL CHARACTER y NATIONAL CHARACTER VARYING). Estos tipos se convierten en tipos CHAR y VARCHAR, respectivamente, y se almacenan en la cantidad especificada de bytes.

Una columna NCHAR sin una especificación de longitud se convierte en una columna CHAR(1).

Una columna NVARCHAR sin una especificación de longitud se convierte en una columna VARCHAR(256).

Tipos TEXT y BPCHAR

Puede crear una tabla de Amazon Redshift con una columna TEXT, pero se convierte a una columna VARCHAR(256) que acepta valores de longitud variable con un máximo de 256 caracteres.

Puede crear una columna de Amazon Redshift con un tipo BPCHAR (carácter con espacio en blanco), que Amazon Redshift convierte a una columna CHAR(256) de longitud fija.

Importancia de los espacios en blancos anteriores y posteriores

Los tipos de datos CHAR y VARCHAR almacenan cadenas de hasta n bytes de longitud. Un intento de almacenar una cadena más grande en una columna de estos tipos genera un error, a menos que los caracteres adicionales sean todos espacios (en blanco), en cuyo caso la cadena se trunca a la longitud máxima. Si la cadena es más corta que la longitud máxima, los valores CHAR se rellenan con espacios en blanco, pero los valores VARCHAR almacenan la cadena sin espacios en blanco.

Los espacios en blanco anteriores o posteriores en valores CHAR no tienen importancia semántica. Se omiten cuando compara dos valores CHAR, no se incluyen en cálculos LENGTH y se eliminan cuando convierte un valor CHAR a otro tipo de cadena.

Los espacios anteriores o posteriores en los valores VARCHAR y CHAR no tienen importancia semántica cuando se comparan valores.

Los cálculos de longitud devuelven la longitud de cadenas de caracteres VARCHAR con espacios anteriores o posteriores incluidos en la longitud. Los espacios anteriores o posteriores no cuentan en la longitud para cadenas de caracteres de longitud fija.

Ejemplos con tipos de caracteres

Instrucción CREATE TABLE

La siguiente instrucción CREATE TABLE demuestra el uso de tipos de datos VARCHAR y CHAR:

```
create table address(
address_id integer,
```

```
address1 varchar(100),
address2 varchar(50),
district varchar(20),
city_name char(20),
state char(2),
postal_code char(5)
);
```

Los siguientes ejemplos utilizan esta tabla.

Espacios en blanco anteriores o posteriores en cadenas de caracteres de longitud variable

Dado que ADDRESS1 es una columna VARCHAR, los espacios en blanco anteriores o posteriores en la segunda dirección insertada no tienen importancia semántica. En otras palabras, estas dos direcciones insertadas coinciden.

```
insert into address(address1) values('9516 Magnolia Boulevard');

insert into address(address1) values('9516 Magnolia Boulevard  ');
```

```
select count(*) from address
where address1='9516 Magnolia Boulevard';

count
-------
2
(1 row)
```

Si la columna ADDRESS1 fuera una columna CHAR y se insertaran los mismo valores, la consulta COUNT(*) reconocería que las cadenas de caracteres son iguales y devolvería 2.

Resultados de la función LENGTH

La función LENGTH reconoce espacios en blancos anteriores o posteriores en columnas VARCHAR:

```
select length(address1) from address;

length
--------
23
25
```

```
(2 rows)
```

Un valor de `Augusta` en la columna CITY_NAME, que es una columna CHAR, siempre devolvería una longitud de 7 caracteres, independientemente de cualquier espacio en blanco anterior o posterior en la cadena de entrada.

Valores que superan la longitud de la columna

Las cadenas de caracteres no se truncan para ajustarse al ancho declarado de la columna:

```
insert into address(city_name) values('City of South San Francisco');
ERROR: value too long for type character(20)
```

Una solución a este problema es formar el valor con el tamaño de la columna:

```
insert into address(city_name)
values('City of South San Francisco'::char(20));
```

En este caso, se cargarían los primeros 20 caracteres de la cadena (`City of South San Fr`) en la columna.

Tipos de fecha y hora

Temas

- Almacenamiento y rangos
- DATE
- TIME
- TIMETZ
- TIMESTAMP
- TIMESTAMPTZ
- Ejemplos con tipos de fecha y hora
- Literales de fecha, hora y marca temporal
- Literales y tipos de datos de intervalo

Los tipos de datos de fecha y hora incluyen DATE, TIME, TIMETZ, TIMESTAMP y TIMESTAMPTZ.

Almacenamiento y rangos

| Nombre | Almacenamiento | Range | Resolución |
|---|---|---|---|
| DATE | 4 bytes | De 4713 a.C. a 294276 d.C. | 1 día |
| TIME | 8 bytes | 00:00:00 a 24:00:00 | 1 microsegundo |
| TIMETZ | 8 bytes | 00:00:00+1459 a 00:00:00+1459 | 1 microsegundo |
| TIMESTAMP | 8 bytes | De 4713 a.C. a 294276 d.C. | 1 microsegundo |
| TIMESTAMPTZ | 8 bytes | De 4713 a.C. a 294276 d.C. | 1 microsegundo |

DATE

Utilice el tipo de datos DATE para almacenar fechas de calendario simples sin marcas temporales.

TIME

TIME es un alias de TIME WITHOUT TIME ZONE.

Utilice el tipo de datos TIME para almacenar la hora del día.

Las columnas TIME almacenan valores con un máximo de seis dígitos de precisión para las fracciones de segundos.

De forma predeterminada, los valores TIME están en hora universal coordinada (UTC) en las tablas de usuario y las tablas de sistema de Amazon Redshift.

TIMETZ

TIMETZ es un alias de TIME WITH TIME ZONE.

Utilice el tipo de datos TIMETZ para almacenar la hora del día con una zona horaria.

Las columnas TIMETZ almacenan valores con un máximo de seis dígitos de precisión para las fracciones de segundos.

De forma predeterminada, los valores TIMETZ están en UTC en las tablas de usuario y las tablas de sistema de Amazon Redshift.

TIMESTAMP

TIMESTAMP es un alias de TIMESTAMP WITHOUT TIME ZONE.

Use el tipo de datos TIMESTAMP para almacenar valores de marca temporal completa que incluyen la fecha y la hora del día.

Las columnas TIMESTAMP almacenan valores con un máximo de seis dígitos de precisión para las fracciones de segundos.

Si inserta una fecha en una columna TIMESTAMP o una fecha con un valor de marca temporal parcial, el valor se convierte de manera implícita a un valor de marca temporal completa. Este valor de marca temporal completo tiene valores predeterminados (00) para las horas, los minutos y los segundos que faltan. Se ignoran los valores de zona horaria en cadenas de entrada.

De forma predeterminada, los valores TIMESTAMP están en UTC en las tablas de usuario y las tablas de sistema de Amazon Redshift.

TIMESTAMPTZ

TIMESTAMPTZ es un alias de TIMESTAMP WITH TIME ZONE.

Use el tipo de datos TIMESTAMPTZ para introducir valores de marca temporal completa que incluyan la fecha, la hora del día y una zona horaria. Cuando un valor de entrada incluye una zona horaria, Amazon Redshift utiliza la zona horaria para convertir el valor a la UTC y almacena el valor de la UTC.

Para ver una lista de los nombres de zonas horarias compatibles, ejecute el siguiente comando.

```
select pg_timezone_names();
```

Para ver una lista de las abreviaturas de zonas horarias compatibles, ejecute el siguiente comando.

```
select pg_timezone_abbrevs();
```

Puede obtener información actualizada acerca de las zonas horarias en la base de datos de zonas horarias de IANA.

La siguiente tabla tiene ejemplos de formatos de zonas horarias.

| Formato | Ejemplo |
| --- | --- |
| dd mon hh:mi:ss yyyy tz | 17 dic 07:37:16 1997 PST |
| mm/dd/yyyy hh:mi:ss.ss tz | 12/17/1997 07:37:16.00 PST |
| mm/dd/yyyy hh:mi:ss.ss tz | 12/17/1997 07:37:16.00 EE. UU./Pacífico |
| yyyy-mm-dd hh:mi:ss+/-tz | 1997-12-17 07:37:16-08 |
| dd.mm.yyyy hh:mi:ss tz | 17.12.1997 07:37:16.00 PST |

Las columnas TIMESTAMPTZ almacenan valores con un máximo de seis dígitos de precisión para las fracciones de segundos.

Si inserta una fecha en una columna TIMESTAMPTZ o una fecha con una marca temporal parcial, el valor se convierte de manera implícita a un valor de marca temporal completa. Este valor de marca temporal completo tiene valores predeterminados (00) para las horas, los minutos y los segundos que faltan.

Los valores TIMESTAMPTZ son UTC en tablas del usuario.

Ejemplos con tipos de fecha y hora

A continuación, encontrará ejemplos de cómo trabajar con tipos de fecha y hora compatibles con Amazon Redshift.

Ejemplos de fecha

Los siguientes ejemplos insertan fechas que tienen diferentes formatos y muestran la salida.

```
create table datetable (start_date date, end_date date);
```

```
insert into datetable values ('2008-06-01','2008-12-31');

insert into datetable values ('Jun 1,2008','20081231');
```

```
select * from datetable order by 1;

start_date |  end_date
```

```
-----------------------
2008-06-01 | 2008-12-31
2008-06-01 | 2008-12-31
```

Si inserta un valor de marca temporal en una columna DATE, se ignora la parte de la hora y solo se carga la fecha.

Ejemplos de tiempo

Los siguientes ejemplos insertan los valores TIME y TIMETZ que tienen diferentes formatos y muestran la salida.

```
create table timetable (start_time time, end_time timetz);
```

```
insert into timetable values ('19:11:19','20:41:19 UTC');
insert into timetable values ('191119', '204119 UTC');
```

```
select * from timetable order by 1;
start_time |  end_time
-----------------------
 19:11:19  | 20:41:19+00
 19:11:19  | 20:41:19+00
```

Ejemplos de marca temporal

Si inserta una fecha en una columna TIMESTAMP o TIMESTAMPTZ, la hora es, por defecto, la medianoche. Por ejemplo, si inserta el literal `20081231`, el valor almacenado es `2008-12-31 00:00:00`.

Para cambiar la zona horaria de la sesión actual, use el comando SET para establecer el parámetro de configuración timezone.

En el siguiente ejemplo se insertan marcas temporales que tienen diferentes formatos y muestran la tabla de resultados.

```
create table tstamp(timeofday timestamp, timeofdaytz timestamptz);

insert into tstamp values('Jun 1,2008  09:59:59', 'Jun 1,2008 09:59:59 EST' );
insert into tstamp values('Dec 31,2008 18:20','Dec 31,2008 18:20');
insert into tstamp values('Jun 1,2008  09:59:59 EST', 'Jun 1,2008 09:59:59');
```

```
SELECT * FROM tstamp;

+--------------------+------------------------+
|      timeofday     |      timeofdaytz       |
+--------------------+------------------------+
| 2008-06-01 09:59:59 | 2008-06-01 14:59:59+00 |
| 2008-12-31 18:20:00 | 2008-12-31 18:20:00+00 |
| 2008-06-01 09:59:59 | 2008-06-01 09:59:59+00 |
+--------------------+------------------------+
```

Literales de fecha, hora y marca temporal

A continuación, se indican las reglas para trabajar con literales de fecha, hora y marca temporal compatibles con Amazon Redshift.

Fechas

Las siguientes fechas de entrada son ejemplos válidos de valores de fecha literales del tipo de dato DATE que puede cargar en tablas de Amazon Redshift. Se supone que el modo predeterminado `MDY` `DateStyle` está en vigor. Este modo significa que el valor del mes precede al valor del día en las cadenas, como `1999-01-08` y `01/02/00`.

> **Note**
>
> Un literal de marca temporal o fecha debe encerrarse entre comillas cuando lo carga a la tabla.

| Fecha de entrada | Fecha completa |
|---|---|
| 8 de enero de 1999 | 8 de enero de 1999 |
| 1999-01-08 | 8 de enero de 1999 |
| 1/8/1999 | 8 de enero de 1999 |
| 01/02/00 | 2 de enero de 2000 |
| 2000-Ene-31 | 31 de enero de 2000 |
| Ene-31-2000 | 31 de enero de 2000 |

| Fecha de entrada | Fecha completa |
| --- | --- |
| 31-Ene-2000 | 31 de enero de 2000 |
| 20080215 | 15 de febrero de 2008 |
| 080215 | 15 de febrero de 2008 |
| 2008.366 | 31 de diciembre de 2008 (la parte de tres dígitos de la fecha debe tener un valor del rango 001-366). |

Times

Las siguientes horas de entrada son ejemplos válidos de valores de hora literales para los tipos de datos TIME y TIMETZ que puede cargar en tablas de Amazon Redshift.

| Horas de entrada | Descripción (de la parte de la hora) |
| --- | --- |
| 04:05:06789 | 4:05 a. m. y 6789 segundos |
| 04:05:06 | 4:05 a. m. y 6 segundos |
| 04:05 | 4:05 a. m. exactamente |
| 040506 | 4:05 a. m. y 6 segundos |
| 04:05 a. m. | 4:05 a. m. exactamente; a. m. es opcional |
| 04:05. p. m. | 4:05 p. m. exactamente; el valor de la hora debe ser menor que 12 |
| 16:05 | 4:05 p. m. exactamente |

Marcas de tiempo

Las siguientes marcas de tiempo de entrada son ejemplos válidos de valores de tiempo literales para los tipos de datos TIMESTAMP y TIMESTAMPTZ que puede cargar en tablas de Amazon Redshift. Todos los literales de fecha válidos pueden combinarse con los siguientes literales de hora.

| Marcas temporales de entrada (fechas y horas concatenadas) | Descripción (de la parte de la hora) |
|---|---|
| 20080215 04:05:06.789 | 4:05 a. m. y 6789 segundos |
| 20080215 04:05:06 | 4:05 a. m. y 6 segundos |
| 20080215 04:05 | 4:05 a. m. exactamente |
| 20080215 040506 | 4:05 a. m. y 6 segundos |
| 20080215 04:05 AM | 4:05 a. m. exactamente; a. m. es opcional |
| 20080215 04:05 PM | 4:05 p. m. exactamente; el valor de la hora debe ser menor que 12 |
| 20080215 16:05 | 4:05 p. m. exactamente |
| 20080215 | Medianoche (por defecto) |

Valores de fecha y hora especiales

Los siguiente valores especiales se pueden usar como literales de fecha y hora y como argumentos para funciones de fecha. Requieren comillas simples y se convierten en valores de marca temporal regulares durante el procesamiento de consultas.

| Valor especial | Descripción |
|---|---|
| `now` | Evalúa la hora de inicio de la transacción actual y devuelve una marca temporal con precisión de microsegundo. |
| `today` | Toma el valor de la fecha adecuada y devuelve una marca temporal con ceros en las partes de la hora. |
| `tomorrow` | Toma el valor de la fecha adecuada y devuelve una marca temporal con ceros en las partes de la hora. |

| Valor especial | Descripción |
| --- | --- |
| `yesterday` | Toma el valor de la fecha adecuada y devuelve una marca temporal con ceros en las partes de la hora. |

En los siguientes ejemplos, se muestra cómo `now` y `today` trabajan en conjunto con la función DATEADD.

```
select dateadd(day,1,'today');

date_add
---------------------
2009-11-17 00:00:00
(1 row)

select dateadd(day,1,'now');

date_add
----------------------------
2009-11-17 10:45:32.021394
(1 row)
```

Literales y tipos de datos de intervalo

Puede usar un tipo de datos de intervalo para almacenar duraciones de tiempo en unidades como `seconds`, `minutes`, `hours`, `days`, `months` y `years`. Los literales y los tipos de datos de intervalo se pueden usar en los cálculos de fecha y hora, por ejemplo, agregar intervalos a fechas y marcas temporales, sumar intervalos y restar un intervalo de una fecha o marca temporal. Los literales de intervalo se pueden usar como valores de entrada para las columnas de tipos de datos de intervalos de una tabla.

Sintaxis del tipo de datos de intervalo

Para especificar un tipo de datos de intervalo para almacenar una duración de tiempo en años y meses:

```
INTERVAL year_to_month_qualifier
```

Para especificar un tipo de datos de intervalo para almacenar una duración en días, horas, minutos y segundos:

```
INTERVAL day_to_second_qualifier [ (fractional_precision) ]
```

Sintaxis de literal de intervalo

Para especificar un literal de intervalo para definir una duración de tiempo en años y meses:

```
INTERVAL quoted-string year_to_month_qualifier
```

Para especificar un literal de intervalo para definir una duración en días, horas, minutos y segundos:

```
INTERVAL quoted-string day_to_second_qualifier [ (fractional_precision) ]
```

Argumentos

quoted-string

Especifica un valor numérico positivo o negativo especificando una cantidad y la unidad de fecha y hora como cadena de entrada. Si quoted-string contiene solo un valor numérico, Amazon Redshift determina las unidades de year_to_month_qualifier o day_to_second_qualifier. Por ejemplo, `'23' MONTH` representa `1 year 11 months`, `'-2' DAY` representa `-2 days 0 hours 0 minutes 0.0 seconds`, `'1-2' MONTH` representa `1 year 2 months` y `'13 day 1 hour 1 minute 1.123 seconds' SECOND` representa `13 days 1 hour 1 minute 1.123 seconds`. Para obtener más información acerca de los formatos de salida de un intervalo, consulte Estilos de intervalo.

year_to_month_qualifier

Especifica el rango del intervalo. Si usa un calificador y crea un intervalo con unidades de tiempo más pequeñas que el calificador, Amazon Redshift corta y descarta las partes más pequeñas del intervalo. Los valores válidos para year_to_month_qualifier son:

- `YEAR`
- `MONTH`
- `YEAR TO MONTH`

day_to_second_qualifier

Especifica el rango del intervalo. Si usa un calificador y crea un intervalo con unidades de tiempo más pequeñas que el calificador, Amazon Redshift corta y descarta las partes más pequeñas del intervalo. Los valores válidos para day_to_second_qualifier son:

- `DAY`
- `HOUR`
- `MINUTE`
- `SECOND`
- `DAY TO HOUR`
- `DAY TO MINUTE`
- `DAY TO SECOND`
- `HOUR TO MINUTE`
- `HOUR TO SECOND`
- `MINUTE TO SECOND`

El resultado del literal INTERVAL se trunca al componente INTERVAL más pequeño especificado. Por ejemplo, al usar un calificador MINUTE, Amazon Redshift descarta las unidades de tiempo inferiores a MINUTE.

```
select INTERVAL '1 day 1 hour 1 minute 1.123 seconds' MINUTE
```

El valor resultante se trunca en `'1 day 01:01:00'`.

fractional_precision

Parámetro opcional que especifica el número de dígitos fraccionales permitidos en el intervalo. El argumento fractional_precision solo se debe especificar si el intervalo contiene SECOND. Por ejemplo, `SECOND(3)` crea un intervalo que permite solo tres dígitos fraccionales, como 1234 segundos. El número máximo de dígitos fraccionales es seis.

La configuración de la sesión `interval_forbid_composite_literals` determina si se devuelve un error cuando se especifica un intervalo con las partes YEAR TO MONTH y DAY TO SECOND. Para obtener más información, consulte interval_forbid_composite_literals.

### Aritmética de intervalos

Puede utilizar valores de intervalo con otros valores de fecha y hora para realizar operaciones aritméticas. En las siguientes tablas se describen las operaciones disponibles y los resultados de tipo de datos de cada operación.

> **Note**
>
> Las operaciones que pueden producir resultados `date` y `timestamp` lo hacen en función de la unidad de tiempo más pequeña implicada en la ecuación. Por ejemplo, cuando se agrega un `interval` a una `date` el resultado es una `date` si es un intervalo YEAR TO MONTH y una marca temporal si es un intervalo DAY TO SECOND.

Las operaciones en las que el primer operando es un `interval` producen los siguientes resultados para el segundo operando dado:

| Operador | Date | Timestamp | Interval | Numérico |
|---|---|---|---|---|
| - | N/A | N/A | Interval | N/A |
| + | Date | Date/Timestamp | Interval | N/A |
| * | N/A | N/A | N/A | Interval |
| / | N/A | N/A | N/A | Interval |

Las operaciones en las que el primer operando es una `date` producen los siguientes resultados para el segundo operando dado:

| Operador | Date | Timestamp | Interval | Numérico |
|---|---|---|---|---|
| - | Numérico | Interval | Date/Timestamp | Date |
| + | N/A | N/A | N/A | N/A |

Las operaciones en las que el primer operando es una `timestamp` producen los siguientes resultados para el segundo operando dado:

| Operador | Date | Timestamp | Interval | Numérico |
|---|---|---|---|---|
| - | Numérico | Interval | Timestamp | Timestamp |
| + | N/A | N/A | N/A | N/A |

Estilos de intervalo

Puede utilizar el comando the section called “SET” de SQL para cambiar el formato de pantalla de salida de los valores de intervalos. Cuando utilice el tipo de datos de intervalo en SQL, conviértalo en texto para ver el estilo de intervalo esperado, por ejemplo, `YEAR TO MONTH::text`. Los valores disponibles para ESTABLECER el valor `IntervalStyle` son:

- `postgres`: sigue el estilo de PostgreSQL. Esta es la opción predeterminada.
- `postgres_verbose`: sigue el estilo detallado de PostgreSQL.
- `sql_standard`: sigue el estilo de literales de intervalo estándar de SQL.

El siguiente comando establece el estilo de intervalo en `sql_standard`.

```
SET IntervalStyle to 'sql_standard';
```

Formato de salida postgres

A continuación, se muestra el formato de salida del estilo de intervalo `postgres`. Cada valor numérico puede ser negativo.

```
'<numeric> <unit> [, <numeric> <unit> ...]'
```

```
select INTERVAL '1-2' YEAR TO MONTH::text

varchar
---------------
1 year 2 mons
```

```
select INTERVAL '1 2:3:4.5678' DAY TO SECOND::text
```

```
varchar
------------------
1 day 02:03:04.5678
```

Formato de salida postgres_verbose

La sintaxis de postgres_verbose es similar a la de postgres, pero las salidas de postgres_verbose también contienen la unidad de tiempo.

```
'[@] <numeric> <unit> [, <numeric> <unit> ...] [direction]'
```

```
select INTERVAL '1-2' YEAR TO MONTH::text

varchar
-----------------
@ 1 year 2 mons
```

```
select INTERVAL '1 2:3:4.5678' DAY TO SECOND::text

varchar
---------------------------
@ 1 day 2 hours 3 mins 4.56 secs
```

Formato de salida sql_standard

Los valores del intervalo de año a mes tienen el siguiente formato. Si se especifica un signo negativo antes del intervalo, eso indica que el intervalo es un valor negativo y se aplica a todo el intervalo.

```
'[-]yy-mm'
```

Los valores del intervalo de día a segundo tienen el siguiente formato.

```
'[-]dd hh:mm:ss.ffffff'
```

```
SELECT INTERVAL '1-2' YEAR TO MONTH::text

varchar
-------
```

```
1-2
```

```
select INTERVAL '1 2:3:4.5678' DAY TO SECOND::text

varchar
---------------
1 2:03:04.5678
```

Ejemplos de tipo de datos de intervalo

En los siguientes ejemplos, se muestra cómo usar tipos de datos INTERVAL con tablas.

```
create table sample_intervals (y2m interval month, h2m interval hour to minute);
insert into sample_intervals values (interval '20' month, interval '2 days
 1:1:1.123456' day to second);
select y2m::text, h2m::text from sample_intervals;


      y2m      |       h2m
---------------+-----------------
 1 year 8 mons | 2 days 01:01:00
```

```
update sample_intervals set y2m = interval '2' year where y2m = interval '1-8' year to
 month;
select * from sample_intervals;

   y2m   |       h2m
---------+-----------------
 2 years | 2 days 01:01:00
```

```
delete from sample_intervals where h2m = interval '2 1:1:0' day to second;
select * from sample_intervals;

 y2m | h2m
-----+-----
```

Ejemplos de literales de intervalo

Los siguientes ejemplos se ejecutan con el estilo de intervalo establecido en `postgres`.

En el siguiente ejemplo, se muestra cómo crear un literal INTERVAL de 1 año.

```
select INTERVAL '1' YEAR

intervaly2m
---------------
1 years 0 mons
```

Si especifica una quoted-string que supere el calificador, las unidades de tiempo restantes se truncan con respecto al intervalo. En el ejemplo siguiente, un intervalo de 13 meses se convierte en 1 año y 1 mes, pero el mes restante se omite debido al calificador YEAR.

```
select INTERVAL '13 months' YEAR

intervaly2m
---------------
1 years 0 mons
```

Si utiliza un calificador inferior a la cadena de intervalos, se incluyen las unidades sobrantes.

```
select INTERVAL '13 months' MONTH

intervaly2m
---------------
1 years 1 mons
```

Al especificar una precisión en el intervalo, se trunca el número de dígitos fraccionarios hasta alcanzar la precisión especificada.

```
select INTERVAL '1.234567' SECOND (3)

intervald2s
--------------------------------
0 days 0 hours 0 mins 1.235 secs
```

Si no especifica una precisión, Amazon Redshift utiliza la precisión máxima de 6.

```
select INTERVAL '1.23456789' SECOND

intervald2s
```

```
-----------------------------------
0 days 0 hours 0 mins 1.234567 secs
```

En el siguiente ejemplo, se muestra cómo crear un intervalo con rangos.

```
select INTERVAL '2:2' MINUTE TO SECOND

intervald2s
-----------------------------
0 days 0 hours 2 mins 2.0 secs
```

Los calificadores dictan las unidades que se especifican. Por ejemplo, aunque en el siguiente ejemplo se utiliza la misma quoted-string de “2:2” que en el ejemplo anterior, Amazon Redshift reconoce que utiliza diferentes unidades de tiempo debido al calificador.

```
select INTERVAL '2:2' HOUR TO MINUTE

intervald2s
-----------------------------
0 days 2 hours 2 mins 0.0 secs
```

También se admiten las abreviaturas y los plurales de cada unidad. Por ejemplo, `5s`, `5 second` y `5 seconds` son intervalos equivalentes. Las unidades admitidas son años, meses, horas, minutos y segundos.

```
select INTERVAL '5s' SECOND

intervald2s
-----------------------------
0 days 0 hours 0 mins 5.0 secs
```

```
select INTERVAL '5 HOURS' HOUR

intervald2s
-----------------------------
0 days 5 hours 0 mins 0.0 secs
```

```
select INTERVAL '5 h' HOUR
```

```
intervald2s
----------------------------
0 days 5 hours 0 mins 0.0 secs
```

Ejemplos de literales de intervalo sin sintaxis de calificador

> **Note**
>
> En los siguientes ejemplos se muestra el uso de un literal de intervalo sin un calificador `YEAR TO MONTH` o `DAY TO SECOND`. Para obtener información sobre el uso del literal de intervalo recomendado con un calificador, consulte Literales y tipos de datos de intervalo.

Use un literal de intervalo para identificar períodos específicos de tiempo, como `12 hours` o `6 months`. Puede usar estos literales de intervalo en condiciones y cálculos que involucran expresiones de fecha y hora.

Un literal de intervalo se expresa como una combinación de la palabra clave INTERVAL con una cantidad numérica y una parte de fecha compatible, por ejemplo, `INTERVAL '7 days'` o `INTERVAL '59 minutes'`. Puede conectar varias cantidades y unidades para formar un intervalo más preciso, por ejemplo: `INTERVAL '7 days, 3 hours, 59 minutes'`. También se admiten abreviaturas y plurales de cada unidad; por ejemplo: `5 s`, `5 second` y `5 seconds` son intervalos equivalentes.

Si no especifica una parte de fecha, el valor de intervalo representa segundos. Puede especificar el valor de cantidad como una fracción (por ejemplo: `0.5 days`).

En los siguientes ejemplos se muestra una serie de cálculos con diferentes valores de intervalo.

A continuación, se agrega 1 segundo a la fecha especificada.

```
select caldate + interval '1 second' as dateplus from date
where caldate='12-31-2008';
dateplus
--------------------
2008-12-31 00:00:01
(1 row)
```

A continuación, se agrega 1 minuto a la fecha especificada.

```
select caldate + interval '1 minute' as dateplus from date
where caldate='12-31-2008';
dateplus
---------------------
2008-12-31 00:01:00
(1 row)
```

A continuación, se agregan 3 horas y 35 minutos a la fecha especificada.

```
select caldate + interval '3 hours, 35 minutes' as dateplus from date
where caldate='12-31-2008';
dateplus
---------------------
2008-12-31 03:35:00
(1 row)
```

A continuación, se agregan 52 semanas a la fecha especificada.

```
select caldate + interval '52 weeks' as dateplus from date
where caldate='12-31-2008';
dateplus
---------------------
2009-12-30 00:00:00
(1 row)
```

A continuación, se agregan 1 semana, 1 hora, 1 minuto y 1 segundo a la fecha especificada.

```
select caldate + interval '1w, 1h, 1m, 1s' as dateplus from date
where caldate='12-31-2008';
dateplus
---------------------
2009-01-07 01:01:01
(1 row)
```

A continuación, se agregan 12 horas (medio día) a la fecha especificada.

```
select caldate + interval '0.5 days' as dateplus from date
where caldate='12-31-2008';
dateplus
---------------------
```

```
2008-12-31 12:00:00
(1 row)
```

Lo siguiente resta 4 meses al 15 de febrero de 2023 y el resultado es 15 de octubre de 2022.

```
select date '2023-02-15' - interval '4 months';

?column?
---------------------
2022-10-15 00:00:00
```

Lo siguiente resta 4 meses al 31 de marzo de 2023 y el resultado es 30 de noviembre de 2022. El cálculo tiene en cuenta el número de días de un mes.

```
select date '2023-03-31' - interval '4 months';

?column?
---------------------
2022-11-30 00:00:00
```

Tipo booleano

Use el tipo de dato BOOLEAN para almacenar valores verdaderos y falsos en una columna de un byte. En la siguiente tabla se describen los tres estados posibles para un valor booleano y los valores literales que generan ese estado. Independientemente de la cadena de entrada, una columna booleana almacena y produce "t" para verdadero y "f" para falso.

| Estado | Valores literales válidos | Almacenamiento |
|---|---|---|
| True | `TRUE 't' 'true' 'y' 'yes' '1'` | 1 byte |
| False | `FALSE 'f' 'false' 'n' 'no' '0'` | 1 byte |
| Unknown | `NULL` | 1 byte |

Puede usar una comparación IS para comprobar un valor booleano solo como un predicado en la cláusula WHERE. No puede usar la comparación IS con un valor booleano en la lista SELECT.

Ejemplos

Podría usar una columna BOOLEAN para almacenar un estado "Activo/Inactivo" para cada cliente en una tabla CUSTOMER.

```
create table customer(
custid int,
active_flag boolean default true);
```

```
insert into customer values(100, default);
```

```
select * from customer;
custid | active_flag
-------+-------------
   100 | t
```

Si no se especifica un valor predeterminado (`true` o `false`) en la instrucción CREATE TABLE, insertar un valor predeterminado significa insertar un valor nulo.

En este ejemplo, la consulta selecciona usuarios de la tabla USERS que les gustan los deportes, pero no el cine:

```
select firstname, lastname, likesports, liketheatre
from users
where likesports is true and liketheatre is false
order by userid limit 10;

firstname |  lastname  | likesports | liketheatre
----------+------------+------------+-------------
Lars      | Ratliff    | t          | f
Mufutau   | Watkins    | t          | f
Scarlett  | Mayer      | t          | f
Shafira   | Glenn      | t          | f
Winifred  | Cherry     | t          | f
Chase     | Lamb       | t          | f
Liberty   | Ellison    | t          | f
Aladdin   | Haney      | t          | f
Tashya    | Michael    | t          | f
```

```
Lucian    | Montgomery | t          | f
(10 rows)
```

El siguiente ejemplo selecciona usuarios de la tabla USERS para los que se desconoce si les gusta el rock.

```
select firstname, lastname, likerock
from users
where likerock is unknown
order by userid limit 10;

firstname | lastname | likerock
----------+----------+----------
Rafael    | Taylor   |
Vladimir  | Humphrey |
Barry     | Roy      |
Tamekah   | Juarez   |
Mufutau   | Watkins  |
Naida     | Calderon |
Anika     | Huff     |
Bruce     | Beck     |
Mallory   | Farrell  |
Scarlett  | Mayer    |
(10 rows)
```

El siguiente ejemplo devuelve un error porque usa una comparación IS en la lista SELECT.

```
select firstname, lastname, likerock is true as "check"
from users
order by userid limit 10;

[Amazon](500310) Invalid operation: Not implemented
```

El siguiente ejemplo es correcto porque usa una comparación igual (=) en la lista SELECT en lugar de la comparación IS.

```
select firstname, lastname, likerock = true as "check"
from users
order by userid limit 10;

firstname | lastname  | check
----------+-----------+------
```

```
Rafael    | Taylor    |
Vladimir  | Humphrey  |
Lars      | Ratliff   | true
Barry     | Roy       |
Reagan    | Hodge     | true
Victor    | Hernandez | true
Tamekah   | Juarez    |
Colton    | Roy       | false
Mufutau   | Watkins   |
Naida     | Calderon  |
```

Tipo HLLSKETCH

Utilice el tipo de datos HLLSKETCH para bocetos de HyperLogLog. Amazon Redshift admite representaciones de boceto de HyperLogLog que son dispersas o densas. Los bocetos comienzan siendo dispersos y cambian a densos cuando el formato denso es más eficiente para minimizar la huella de memoria que se utiliza.

Amazon Redshift convierte automáticamente un boceto disperso de HyperLogLog al importar, exportar o imprimir bocetos en el siguiente formato JSON.

```
{"logm":15,"sparse":{"indices":[4878,9559,14523],"values":[1,2,1]}}
```

Amazon Redshift utiliza una representación de cadena en formato Base64 para representar un boceto denso de HyperLogLog.

Amazon Redshift utiliza la siguiente representación de cadena en formato Base64 para representar un boceto denso de HyperLogLog.

```
"ABAABA..."
```

El tamaño máximo de un objeto HLLSKETCH es de 24 580 bytes cuando se utiliza en compresión sin formato.

Tipo SUPER

Utilice el tipo de datos SUPER para almacenar datos o documentos semiestructurados como valores.

Los datos semiestructurados no se ajustan a la estructura rígida y tabular del modelo de datos relacionales utilizado en las bases de datos SQL. Contiene etiquetas que referencian entidades

distintas dentro de los datos. Pueden contener valores complejos, como matrices, estructuras anidadas y otras estructuras complejas, que están asociadas a formatos de serialización, como JSON. El tipo de datos SUPER es un conjunto de valores de estructura y matriz sin esquema que abarcan todos los demás tipos escalares de Amazon Redshift.

El tipo de datos SUPER admite hasta 16 MB de datos para un objeto SUPER individual. Para obtener más información sobre el tipo de datos SUPER, incluidos ejemplos de su implementación en una tabla, consulte Datos semiestructurados en Amazon Redshift.

Solo se puede hacer la ingesta de objetos SUPER de más de 1 MB a partir de los siguientes formatos de archivo:

- Parquet
- JSON
- TEXT
- CSV

El tipo de datos SUPER presenta las siguientes propiedades:

- un valor escalar de Amazon Redshift:
  - un nulo
  - un booleano
  - un número, como smallint, entero, bigint, decimal o de coma flotante (como float4 o float8)
  - un valor de cadena, como varchar o char
- un valor complejo:
  - una matriz de valores, incluidos los escalares o los complejos
  - una estructura, también conocida como tupla u objeto, que es un mapa de nombres y valores de atributos (escalar o complejo)

Cualquiera de los dos tipos de valores complejos contiene sus propios valores escalares o complejos sin ninguna restricción de regularidad.

El tipo de datos SUPER admite la persistencia de datos semiestructurados en un formato sin esquema. Aunque el modelo de datos jerárquico puede cambiar, las versiones antiguas de los datos pueden coexistir en la misma columna SUPER.

Amazon Redshift utiliza PartiQL para habilitar la navegación en matrices y estructuras. Amazon Redshift también utiliza la sintaxis de PartiQL para iterar sobre las matrices SUPER. Para obtener más información, consulte Navegación y Consultas de eliminación del anidamiento.

Amazon Redshift utiliza la escritura dinámica para procesar datos SUPER sin esquema sin necesidad de declarar los tipos de datos antes de utilizarlos en la consulta. Para obtener más información, consulte Escritura dinámica.

Puede aplicar políticas de enmascaramiento dinámico de datos a los valores `scalar` de las rutas de las columnas de tipo SUPER. Para obtener más información sobre el enmascaramiento dinámico de datos, consulte Enmascaramiento de datos dinámico. Para obtener más información acerca del uso del enmascaramiento dinámico de datos con el tipo de datos SUPER, consulte Uso del enmascaramiento dinámico de datos con rutas de tipos de datos SUPER.

Tipo VARBYTE

Utilice una columna VARBYTE, VARBINARY o VARYING BINARY para almacenar valores binarios de longitud variable con un límite fijo.

```
varbyte [ (n) ]
```

El número máximo de bytes (n) puede variar de 1 a 16 777 216. El valor predeterminado es 64 000.

Algunos ejemplos en los que podría ser conveniente utilizar un tipo de datos VARBYTE son los siguientes:

- Puede unir tablas en columnas VARBYTE.
- Puede crear vistas materializadas que contengan columnas VARBYTE. Se admite la actualización progresiva de las vistas materializadas que contienen columnas VARBYTE. Sin embargo, a diferencia de las funciones agregadas COUNT, MIN y MAX y GROUP BY en columnas VARBYTE, no son compatibles con la actualización progresiva.

Para garantizar que todos los bytes sean caracteres imprimibles, Amazon Redshift utiliza el formato hexadecimal cuando muestra los valores VARBYTE. Por ejemplo, el siguiente SQL convierte la cadena hexadecimal 6162 en un valor binario. Aunque el valor devuelto es un valor binario, los resultados aparecen como hexadecimal 6162.

```
select from_hex('6162');
```

```
 from_hex
----------
 6162
```

Amazon Redshift es compatible con la conversión entre VARBYTE y los siguientes tipos de datos:

- CHAR
- VARCHAR
- SMALLINT
- INTEGER
- BIGINT

Cuando se hace una conversión con CHAR y VARCHAR, se utiliza el formato UTF-8. Para obtener más información acerca del formato UTF-8, consulte TO_VARBYTE. Cuando se hace una conversión desde SMALLINT, INTEGER y BIGINT, se mantiene el número de bytes del tipo de datos original. Es decir, dos bytes para SMALLINT, cuatro bytes para INTEGER y ocho bytes para BIGINT.

La siguiente instrucción SQL convierte un string VARCHAR en un VARBYTE. Aunque el valor devuelto es un valor binario, los resultados aparecen como hexadecimal 616263.

```
select 'abc'::varbyte;

 varbyte
---------
 616263
```

La siguiente instrucción SQL convierte un valor CHAR de una columna e un VARBYTE. En este ejemplo, se crea una tabla con una columna CHAR (10) (c), se insertan valores de caracteres más cortos que la longitud de 10. La conversión obtenida completa el resultado con un espacio de caracteres (hex “20”) hasta el tamaño de columna definido. Aunque el valor devuelto es un valor binario, los resultados se muestran como hexadecimal.

```
create table t (c char(10));
insert into t values ('aa'), ('abc');
select c::varbyte from t;
          c
----------------------
 6161202020202020202020
 6162632020202020202020
```

La siguiente instrucción SQL convierte una cadena SMALLINT en un VARBYTE. Aunque el valor devuelto es un valor binario, los resultados aparecen como hexadecimal `0005`, es decir, dos bytes o cuatro caracteres hexadecimales.

```
select 5::smallint::varbyte;

 varbyte
---------
 0005
```

La siguiente instrucción SQL convierte un INTEGER en un VARBYTE. Aunque el valor devuelto es un valor binario, los resultados aparecen como hexadecimal `00000005`, es decir, cuatro bytes u ocho caracteres hexadecimales.

```
select 5::int::varbyte;

 varbyte
----------
 00000005
```

La siguiente instrucción SQL convierte un BIGINT en un VARBYTE. Aunque el valor devuelto es un valor binario, los resultados aparecen como hexadecimal `0000000000000005`, es decir, ocho bytes o 16 caracteres hexadecimales.

```
select 5::bigint::varbyte;

     varbyte
------------------
 0000000000000005
```

Las características de Amazon Redshift compatibles con el tipo de datos VARBYTE son las siguientes:

- Operadores VARBYTE
- CONCAT
- LEN
- Función LENGTH

- OCTET_LENGTH
- Función SUBSTRING
- FROM_HEX
- TO_HEX
- FROM_VARBYTE
- TO_VARBYTE
- GETBIT
- Carga de una columna de tipo de datos VARBYTE
- Descarga de una columna de tipo de datos VARBYTE

Limitaciones a la hora de utilizar el tipo de datos VARBYTE con Amazon Redshift

A continuación, se indican las limitaciones a la hora de utilizar el tipo de datos VARBYTE con Amazon Redshift:

- Amazon Redshift Spectrum solo admite el tipo de datos VARBYTE para archivos Parquet y ORC.
- El editor de consultas de Amazon Redshift y el de Amazon Redshift v2 aún no son totalmente compatibles con el tipo de datos VARBYTE. Por ello, se debe utilizar un cliente SQL diferente cuando trabaje con expresiones VARBYTE.

  Una solución para utilizar el editor de consultas es que, si la longitud de sus datos es inferior a 64 KB y el contenido es un UTF-8 válido, se pueden convertir los valores VARBYTE en VARCHAR, por ejemplo:

  ```
  select to_varbyte('6162', 'hex')::varchar;
  ```
- No se pueden utilizar tipos de datos VARBYTE con funciones definidas por el usuario (UDF) de Python o Lambda.
- No se puede crear una columna HLLSKETCH a partir de una columna VARBYTE ni utilizar APPROXIMATE COUNT DISTINCT en una columna VARBYTE.
- Solo se puede hacer la ingesta de valores VARBYTE de más de 1 MB a partir de los siguientes formatos de archivo:
  - Parquet
  - Texto
  - Valores separados por comas (CSV)

## Conversión y compatibilidad de tipos

A continuación, puede encontrar un análisis del funcionamiento de las reglas de conversión de tipos y la compatibilidad de tipos de datos en Amazon Redshift.

### Compatibilidad

La vinculación de tipos de datos y la vinculación de valores literales y constantes con tipos de datos ocurren durante varias operaciones de la base de datos, incluidas las siguientes:

- Operaciones de Data Manipulation Language (DML, Lenguaje de manipulación de datos) en tablas
- Consultas UNION, INTERSECT y EXCEPT
- Expresiones CASE
- Evaluación de predicados, como LIKE e IN
- La evaluación de funciones SQL que realizan comparaciones o extracciones de datos.
- Comparaciones con operadores matemáticos

Los resultados de estas operaciones dependen de las reglas de conversión de tipos y la compatibilidad de tipos de datos. Compatibilidad implica que no siempre se requiere una correspondencia de uno a uno entre un determinado valor y un determinado tipo de datos. Dado que algunos tipos de datos son compatibles, es posible una conversión implícita o coerción (para obtener más información, consulte Tipos de conversiones implícitas). Cuando los tipos de datos no son compatibles, a menudo puede convertir un valor de un tipo de datos a otro al utilizar la función de conversión explícita.

### Reglas generales de conversión y compatibilidad

Tenga en cuenta las siguientes reglas de conversión y compatibilidad:

- En general, los tipos de datos que caen en la misma categoría (como diferentes tipos de datos numéricos) son compatibles y se pueden convertir implícitamente.

  Por ejemplo, con la conversión implícita puede insertar un valor decimal en una columna de enteros. El decimal se redondea para producir un número entero. O bien, puede extraer un valor numérico, como `2008`, de una fecha e insertar ese valor en una columna de enteros.
- Los tipos de datos numéricos implementan condiciones de desbordamiento que ocurren cuando intenta insertar valores fuera de rango. Por ejemplo, un valor decimal con una precisión de 5 no encaja en una columna decimal que se definió con una precisión de 4. Nunca se trunca un

entero o la parte completa de un decimal; sin embargo, la parte fraccional de un decimal se puede redondear para abajo o arriba, según sea adecuado. Sin embargo, no se redondean los resultados de formas explícitas de los valores seleccionados de tablas.

- Diferentes tipos de cadenas de caracteres son compatibles; las cadenas de la columna VARCHAR que contienen datos de un byte y las cadenas de la columna CHAR se pueden comparar y son convertibles de manera implícita. No se pueden comparar las cadenas VARCHAR que contienen datos multibyte. También puede convertir una cadena de caracteres a una fecha, una hora, una marca temporal o un valor numérico si la cadena es un valor literal adecuado; se omitirán los espacios en blanco anteriores o posteriores. En cambio, puede convertir una fecha, una hora, una marca temporal o un valor numérico a una cadena de caracteres de longitud fija o variable.

  > **Note**
  >
  > Una cadena de caracteres que desea transformar a un tipo numérico debe contener una representación de carácter de un número. Por ejemplo, puede transformar las cadenas `'1.0'` o `'5.9'` a valores decimales, pero no puede transformar la cadena `'ABC'` a tipos numéricos.

- Si compara valores de tipo DECIMAL con cadenas de caracteres, Amazon Redshift intentará convertir la cadena de caracteres en un valor DECIMAL. Al comparar todos los demás valores numéricos con cadenas de caracteres, los valores numéricos se convierten en cadenas de caracteres. Para aplicar la conversión opuesta (por ejemplo, convertir cadenas de caracteres en números enteros o convertir valores de tipo DECIMAL en cadenas de caracteres), utilice una función explícita, como CAST.
- Para convertir valores DECIMAL o NUMERIC de 64 bits a una precisión más grande, debe usar una función de conversión explícita, como las funciones CAST o CONVERT.
- Al convertir DATE o TIMESTAMP a TIMESTAMPTZ, o bien, convertir TIME a TIMETZ, la zona horaria se establece como la zona de la sesión actual. Por defecto, la zona horaria de la sesión es UTC. Para obtener más información acerca de cómo configurar la zona horaria de la sesión, consulte timezone.
- De manera similar, TIMESTAMPTZ se convierte a DATE, TIME o TIMESTAMP en función de la zona horaria de la sesión actual. Por defecto, la zona horaria de la sesión es UTC. Después de la conversión, se elimina la información de la zona horaria.
- Las cadenas de caracteres que representan una marca temporal con zona horaria especificada se convierten a TIMESTAMPTZ con la zona horaria de la sesión actual, que es la UTC de forma predeterminada. Del mismo modo, las cadenas de caracteres que representan una hora con zona

horaria especificada se convierten a TIMETZ con la zona horaria de la sesión actual, que es la UTC de forma predeterminada.

### Tipos de conversiones implícitas

Existen dos tipos de conversiones implícitas:

- Conversiones implícitas en asignaciones, como establecer valores en comandos INSERT o UPDATE.
- Conversiones implícitas en expresiones, como realizar comparaciones en la cláusula WHERE.

En la siguiente tabla se enumeran los tipos de datos que pueden convertirse implícitamente en asignaciones o expresiones. También puede usar una función de conversión explícita para realizar estas conversiones.

| Del tipo | Al tipo |
|---|---|
| BIGINT (INT8) | BOOLEAN |
| | CHAR |
| | DECIMAL (NUMERIC) |
| | DOUBLE PRECISION (FLOAT8) |
| | INTEGER (INT, INT4) |
| | REAL (FLOAT4) |
| | SMALLINT (INT2) |
| | VARCHAR |
| CHAR | VARCHAR |
| DATE | CHAR |
| | VARCHAR |
| | TIMESTAMP |

| Del tipo | Al tipo |
| --- | --- |
| | TIMESTAMPTZ |
| DECIMAL (NUMERIC) | BIGINT (INT8) |
| | CHAR |
| | DOUBLE PRECISION (FLOAT8) |
| | INTEGER (INT, INT4) |
| | REAL (FLOAT4) |
| | SMALLINT (INT2) |
| | VARCHAR |
| DOUBLE PRECISION (FLOAT8) | BIGINT (INT8) |
| | CHAR |
| | DECIMAL (NUMERIC) |
| | INTEGER (INT, INT4) |
| | REAL (FLOAT4) |
| | SMALLINT (INT2) |
| | VARCHAR |
| INTEGER (INT, INT4) | BIGINT (INT8) |
| | BOOLEAN |
| | CHAR |
| | DECIMAL (NUMERIC) |
| | DOUBLE PRECISION (FLOAT8) |

| Del tipo | Al tipo |
| --- | --- |
| | REAL (FLOAT4) |
| | SMALLINT (INT2) |
| | VARCHAR |
| REAL (FLOAT4) | BIGINT (INT8) |
| | CHAR |
| | DECIMAL (NUMERIC) |
| | INTEGER (INT, INT4) |
| | SMALLINT (INT2) |
| | VARCHAR |
| SMALLINT (INT2) | BIGINT (INT8) |
| | BOOLEAN |
| | CHAR |
| | DECIMAL (NUMERIC) |
| | DOUBLE PRECISION (FLOAT8) |
| | INTEGER (INT, INT4) |
| | REAL (FLOAT4) |
| | VARCHAR |
| TIMESTAMP | CHAR |
| | DATE |
| | VARCHAR |

| Del tipo | Al tipo |
|---|---|
| | TIMESTAMPTZ |
| | TIME |
| TIMESTAMPTZ | CHAR |
| | DATE |
| | VARCHAR |
| | TIMESTAMP |
| | TIMETZ |
| TIME | VARCHAR |
| | TIMETZ |
| | INTERVALO DÍA A SEGUNDO |
| TIMETZ | VARCHAR |
| | TIME |
| GEOMETRY | GEOGRAPHY |
| GEOGRAPHY | GEOMETRY |

**Note**

Las conversiones implícitas entre TIMESTAMPTZ, TIMESTAMP, DATE, TIME, TIMETZ o cadenas de caracteres utilizan la zona horaria de la sesión actual. Para obtener más información acerca de configurar la zona horaria actual, consulte timezone.
Los tipos de datos GEOMETRY y GEOGRAPHY no se pueden convertir de forma implícita en ningún otro tipo de datos, excepto entre ellos. Para obtener más información, consulte Función CAST.

El tipo de datos VARBYTE no se puede convertir de forma implícita en otros tipos de datos. Para obtener más información, consulte Función CAST.

Uso de la escritura dinámica para el tipo de datos SUPER

Amazon Redshift utiliza la escritura dinámica para procesar datos SUPER sin esquema sin necesidad de declarar los tipos de datos antes de utilizarlos en la consulta. La escritura dinámica utiliza los resultados de navegar por columnas de datos SUPER sin tener que convertirlos explícitamente en tipos de Amazon Redshift. Para obtener más información acerca del uso de la escritura dinámica para el tipo de datos SUPER, consulte Escritura dinámica.

Puede convertir valores SUPER en otros tipos de datos y a partir de ellos con algunas excepciones. Para obtener más información, consulte Limitaciones.

### Secuencias de intercalación

Amazon Redshift no admite secuencias de intercalación definidas por el usuario o específicas de una ubicación. En general, los resultados de cualquier predicado en cualquier contexto pueden verse afectados por la falta de reglas específicas de un lugar para ordenar y comparar valores de datos. Por ejemplo, las expresiones ORDER BY y las funciones como MIN, MAX y RANK devuelven resultados basados en la ordenación UTF8 binaria de los datos que no tiene en cuenta los caracteres específicos de un lugar.

## Expressions

Temas

- Expresiones simples
- Expresiones compuestas
- Listas de expresiones
- Subconsultas escalares
- Expresiones de función

Una expresión es una combinación de uno o más valores, operadores o funciones que toman el valor de un valor. El tipo de datos de una expresión es, por lo general, el de sus componentes.

## Expresiones simples

Una expresión simple es una de los siguientes:

- Un valor literal o una constante
- Un nombre de columna o una referencia de columna
- Una función escalar
- Una función de agregación (fija)
- Una función de ventana
- Una subconsulta escalar

Algunos ejemplos de expresiones simples incluyen los siguientes:

```
5+12
dateid
sales.qtysold * 100
sqrt (4)
max (qtysold)
(select max (qtysold) from sales)
```

## Expresiones compuestas

Una expresión compuesta es una serie de expresiones simples combinadas por operadores aritméticos. Una expresión simple utilizada en una expresión compuesta debe devolver un valor numérico.

Sintaxis

```
expression
operator
expression | (compound_expression)
```

Argumentos

expresión

Una expresión simple que toma el valor de un valor.

operador

Una expresión aritmética compuesta que se puede construir a través de los siguientes operadores, en este orden de prioridad:

- ( ) : paréntesis para controlar el orden de evaluación
- + , - : operador/signo de positivo y negativo
- ^ , |/ , ||/ : potencia, raíz cuadrada, raíz cúbica
- * , / , % : operadores de multiplicación, división y módulo
- @ : valor absoluto
- + , - : suma y resta
- Operadores & , |, #, ~, **<<**, **>>** : AND, OR, NOT, cambio bit a bit hacia la derecha y cambio bit a bit hacia la izquierda
- ||: concatenación

(compound_expression) (expresión_compuesta)

Las expresiones compuestas pueden anidarse a través del uso de paréntesis.

Ejemplos

Algunos ejemplos de expresiones compuestas incluyen los siguientes.

```
('SMITH' || 'JONES')
sum(x) / y
sqrt(256) * avg(column)
rank() over (order by qtysold) / 100
(select (pricepaid - commission) from sales where dateid = 1882) * (qtysold)
```

Algunas funciones también pueden anidarse dentro de otras funciones. Por ejemplo, cualquier función escalar puede anidarse dentro de otra función escalar. En el siguiente ejemplo se devuelve la suma de los valores absolutos de un conjunto de números:

```
sum(abs(qtysold))
```

Las funciones de ventana no pueden utilizarse como argumentos para las funciones de agregación o para otras funciones de ventana. La siguiente expresión devolvería un error:

```
avg(rank() over (order by qtysold))
```

Las funciones de ventana pueden tener una función de agregación anidada. La siguiente expresión suma conjuntos de valores y, luego, los clasifica:

```
rank() over (order by sum(qtysold))
```

## Listas de expresiones

Una lista de expresiones es una combinación de expresiones y puede aparecer en condiciones de membresía y comparación (cláusulas WHERE) y en cláusulas GROUP BY.

### Sintaxis

```
expression , expression , ... | (expression, expression, ...)
```

### Argumentos

expresión

Una expresión simple que toma el valor de un valor. Una lista de expresiones puede contener una o más expresiones separadas por comas o uno o más conjuntos de expresiones separados por comas. Cuando hay varios conjuntos de expresiones, cada conjunto debe contener la misma cantidad de expresiones y estar separado por paréntesis. La cantidad de expresiones en cada conjunto debe coincidir con la cantidad de expresiones antes del operador en la condición.

### Ejemplos

Los siguientes ejemplos son ejemplos de listas de expresión en condiciones:

```
(1, 5, 10)
('THESE', 'ARE', 'STRINGS')
(('one', 'two', 'three'), ('blue', 'yellow', 'green'))
```

La cantidad de expresiones en cada conjunto debe coincidir con la primera parte de la instrucción:

```
select * from venue
where (venuecity, venuestate) in (('Miami', 'FL'), ('Tampa', 'FL'))
```

```
order by venueid;

venueid |        venuename        | venuecity | venuestate | venueseats
---------+-------------------------+-----------+------------+------------
28 | American Airlines Arena | Miami     | FL         |          0
54 | St. Pete Times Forum    | Tampa     | FL         |          0
91 | Raymond James Stadium   | Tampa     | FL         |      65647
(3 rows)
```

## Subconsultas escalares

Una subconsulta escalar es una consulta SELECT regular entre paréntesis que devuelve exactamente un valor: una fila con una columna. La consulta se ejecuta y el valor devuelto se utiliza en la consulta externa. Si la subconsulta devuelve cero filas, el valor de la expresión de subconsulta es nulo. Si devuelve más de una fila, Amazon Redshift devuelve un error. La subconsulta puede referirse a variables de la consulta principal, que actuarán como constantes durante una invocación de la subconsulta.

Puede usar las subconsultas escalares en la mayoría de instrucciones que requiere una expresión. Las subconsultas escalares no son expresiones válidas en los siguientes casos:

- Como valores predeterminados para expresiones
- En cláusulas GROUP BY y HAVING

### Ejemplo

La siguiente subconsulta calcula el precio promedio pagado por venta en todo el 2008, luego la consulta externa utiliza ese valor en el resultado para compararlo con el precio promedio por venta por trimestre:

```
select qtr, avg(pricepaid) as avg_saleprice_per_qtr,
(select avg(pricepaid)
from sales join date on sales.dateid=date.dateid
where year = 2008) as avg_saleprice_yearly
from sales join date on sales.dateid=date.dateid
where year = 2008
group by qtr
order by qtr;
qtr  | avg_saleprice_per_qtr | avg_saleprice_yearly
-------+-----------------------+----------------------
1     |                647.64 |               642.28
```

```
2     |               646.86 |               642.28
3     |               636.79 |               642.28
4     |               638.26 |               642.28
(4 rows)
```

### Expresiones de función

#### Sintaxis

Cualquier incorporación puede utilizarse como expresión. La sintaxis de una llamada de función es el nombre de una función seguido de su lista de argumentos en paréntesis.

```
function ( [expression [, expression...]] )
```

#### Argumentos

función

Cualquier función integrada. Si desea ver algunas funciones de ejemplo, consulte Referencia de funciones SQL.

expresión

Cualquier expresión que coincida con el tipo de datos y el conteo de parámetros esperado por la función.

#### Ejemplos

```
abs (variable)
select avg (qtysold + 3) from sales;
select dateadd (day,30,caldate) as plus30days from date;
```

## Condiciones

#### Temas

- Sintaxis
- Condición de comparación
- Condiciones lógicas
- Condiciones de coincidencia de patrones

- Condición de rango BETWEEN
- Condición nula
- Condición EXISTS
- Condición IN

Una condición es una instrucción de una o más expresiones y operadores lógicos que toma el valor de verdadero, falso o desconocido. A las condiciones a veces se las denomina predicados.

> **Note**
>
> Todas las comparaciones de cadenas y coincidencias del patrón LIKE distinguen entre mayúsculas y minúsculas. Por ejemplo, "A" y "a" no coinciden. Sin embargo, puede hacer una coincidencia de patrones que no distinga entre mayúsculas y minúsculas al utilizar el predicado ILIKE.

## Sintaxis

```
comparison_condition
| logical_condition
| range_condition
| pattern_matching_condition
| null_condition
| EXISTS_condition
| IN_condition
```

## Condición de comparación

Las condiciones de comparación indican relaciones lógicas entre dos valores. Todas las condiciones de comparación son operadores binarios con un tipo devuelto booleano. Amazon Redshift admite los operadores de comparación que se describen en la tabla siguiente:

| Operador | Sintaxis | Descripción |
|---|---|---|
| < | a < b | El valor a es inferior al valor b. |
| > | a > b | El valor a es superior al valor b. |

| Operador | Sintaxis | Descripción |
|---|---|---|
| <= | `a <= b` | El valor a es inferior o igual al valor b. |
| >= | `a >= b` | El valor a es superior o igual al valor b. |
| = | `a = b` | El valor a es igual al valor b. |
| <> o != | `a <> b or a != b` | El valor a no es igual al valor b. |
| ANY \| SOME | `a = ANY(subquery)` | El valor a es igual a cualquier valor que la subconsulta devuelve. |
| ALL | `a <> ALL or != ALL (subquery))` | El valor a no es igual a cualquier valor que la subconsulta devuelve. |
| IS TRUE \| FALSE \| UNKNOWN | `a IS TRUE` | El valor a es un valor TRUE booleano. |

Notas de uso

= ANY | SOME

Las palabras clave ANY y SOME son sinónimos de la condición IN y devuelven true si la comparación es verdadera para al menos un valor devuelto por una subconsulta que brinda uno o más valores. Amazon Redshift solo admite la condición = (igual) para ANY y SOME. No se admiten las condiciones de desigualdad.

> **Note**
>
> Se admite el predicado ALL.

<> ALL

La palabra clave ALL es sinónimo de NOT IN (consulte la condición Condición IN) y devuelve true si la expresión no se incluye en los resultados de una subconsulta. Amazon Redshift solo admite la condición <> o != (no es igual) para ALL. No se admiten otras condiciones de comparación.

### IS TRUE/FALSE/UNKNOWN

Los valores distintos de cero tienen un valor de TRUE, 0 tiene un valor de FALSE, y los valores nulos tienen un valor de UNKNOWN. Consulte el tipo de datos Tipo booleano.

### Ejemplos

A continuación se muestran algunos ejemplos sencillos de condiciones de comparación:

```
a = 5
a < b
min(x) >= 5
qtysold = any (select qtysold from sales where dateid = 1882
```

La siguiente consulta devuelve lugares con más de 10 000 asientos de la tabla VENUE:

```
select venueid, venuename, venueseats from venue
where venueseats > 10000
order by venueseats desc;

venueid |           venuename            | venueseats
---------+--------------------------------+------------
83 | FedExField                     |      91704
 6 | New York Giants Stadium        |      80242
79 | Arrowhead Stadium              |      79451
78 | INVESCO Field                  |      76125
69 | Dolphin Stadium                |      74916
67 | Ralph Wilson Stadium           |      73967
76 | Jacksonville Municipal Stadium |      73800
89 | Bank of America Stadium        |      73298
72 | Cleveland Browns Stadium       |      73200
86 | Lambeau Field                  |      72922
...
(57 rows)
```

Este ejemplo selecciona los usuarios (USERID) de la tabla USERS que les gusta el rock:

```
select userid from users where likerock = 't' order by 1 limit 5;

userid
--------
3
```

```
5
6
13
16
(5 rows)
```

Este ejemplo selecciona los usuarios (USERID) de la tabla USERS para los que se desconoce si les gusta el rock:

```
select firstname, lastname, likerock
from users
where likerock is unknown
order by userid limit 10;

firstname | lastname | likerock
----------+----------+----------
Rafael    | Taylor   |
Vladimir  | Humphrey |
Barry     | Roy      |
Tamekah   | Juarez   |
Mufutau   | Watkins  |
Naida     | Calderon |
Anika     | Huff     |
Bruce     | Beck     |
Mallory   | Farrell  |
Scarlett  | Mayer    |
(10 rows
```

Ejemplos con una columna TIME

La siguiente tabla de ejemplo, TIME_TEST, tiene una columna TIME_VAL (tipo TIME) con tres valores insertados.

```
select time_val from time_test;

time_val
--------------------
20:00:00
00:00:00.5550
00:58:00
```

En el siguiente ejemplo, se extraen las horas de cada timetz_val.

```
select time_val from time_test where time_val < '3:00';
   time_val
---------------
 00:00:00.5550
 00:58:00
```

En el siguiente ejemplo, se comparan dos literales de tiempo.

```
select time '18:25:33.123456' = time '18:25:33.123456';
 ?column?
----------
 t
```

Ejemplos con una columna TIMETZ

La siguiente tabla de ejemplo, TIMETZ_TEST, tiene una columna TIMETZ_VAL (tipo TIMETZ) con tres valores insertados.

```
select timetz_val from timetz_test;

timetz_val
------------------
04:00:00+00
00:00:00.5550+00
05:58:00+00
```

En el siguiente ejemplo, se seleccionan solo los valores TIMETZ menores que `3:00:00 UTC`. La comparación se realiza después de convertir el valor a la UTC.

```
select timetz_val from timetz_test where timetz_val < '3:00:00 UTC';

   timetz_val
---------------
 00:00:00.5550+00
```

En el siguiente ejemplo, se comparan dos literales TIMETZ. Para la comparación, se ignora la zona horaria.

```
select time '18:25:33.123456 PST' < time '19:25:33.123456 EST';
```

```
 ?column?
----------
 t
```

## Condiciones lógicas

Las condiciones lógicas combinan el resultado de dos condiciones para producir un único resultado. Todas las condiciones lógicas son operadores binarios con un tipo devuelto booleano.

Sintaxis

```
expression
{ AND | OR }
expression
NOT expression
```

Las condiciones lógicas utilizan un lógico booleano de tres valores donde el valor nulo representa una relación desconocida. En la siguiente tabla se describen los resultados de condiciones lógicas, donde `E1` y `E2` representan expresiones:

| E1 | E2 | E1 AND E2 | E1 OR E2 | NOT E2 |
|---|---|---|---|---|
| TRUE | TRUE | TRUE | TRUE | FALSO |
| TRUE | FALSO | FALSO | TRUE | TRUE |
| TRUE | UNKNOWN | UNKNOWN | TRUE | UNKNOWN |
| FALSO | TRUE | FALSO | TRUE | |
| FALSO | FALSO | FALSO | FALSO | |
| FALSO | UNKNOWN | FALSO | UNKNOWN | |
| UNKNOWN | TRUE | UNKNOWN | TRUE | |
| UNKNOWN | FALSO | FALSO | UNKNOWN | |
| UNKNOWN | UNKNOWN | UNKNOWN | UNKNOWN | |

El operador NOT se evalúa antes de AND, y el operador AND se evalúa antes del operador OR. Cualquier paréntesis utilizado puede invalidar este orden de evaluación predeterminado.

Ejemplos

En el siguiente ejemplo se devuelve USERID y USERNAME de la tabla USERS donde al usuario le gusta Las Vegas y los deportes:

```
select userid, username from users
where likevegas = 1 and likesports = 1
order by userid;

userid | username
--------+----------
1 | JSG99FHE
67 | TWU10MZT
87 | DUF19VXU
92 | HYP36WEQ
109 | FPL38HZK
120 | DMJ24GUZ
123 | QZR22XGQ
130 | ZQC82ALK
133 | LBN45WCH
144 | UCX04JKN
165 | TEY68OEB
169 | AYQ83HGO
184 | TVX65AZX
...
(2128 rows)
```

En el siguiente ejemplo se devuelve el USERID y USERNAME de la tabla USERS donde al usuario le gusta Las Vegas o los deportes, o ambos. Esta consulta devuelve todos los resultados del ejemplo anterior además de los usuarios que solo les gustan Las Vegas o los deportes.

```
select userid, username from users
where likevegas = 1 or likesports = 1
order by userid;

userid | username
--------+----------
1 | JSG99FHE
2 | PGL08LJI
```

```
3 | IFT66TXU
5 | AEB55QTM
6 | NDQ15VBM
9 | MSD36KVR
10 | WKW41AIW
13 | QTF33MCG
15 | OWU78MTR
16 | ZMG93CDD
22 | RHT62AGI
27 | KOY02CVE
29 | HUH27PKK
...
(18968 rows)
```

La siguiente consulta usa paréntesis alrededor de la condición OR para encontrar lugares en Nueva York o California donde se realizó Macbeth:

```
select distinct venuename, venuecity
from venue join event on venue.venueid=event.venueid
where (venuestate = 'NY' or venuestate = 'CA') and eventname='Macbeth'
order by 2,1;

venuename                |   venuecity
-----------------------------------------+---------------
Geffen Playhouse                         | Los Angeles
Greek Theatre                            | Los Angeles
Royce Hall                               | Los Angeles
American Airlines Theatre                | New York City
August Wilson Theatre                    | New York City
Belasco Theatre                          | New York City
Bernard B. Jacobs Theatre                | New York City
...
```

Eliminar los paréntesis en este ejemplo cambia la lógica y los resultados de la consulta.

En el siguiente ejemplo se usa el operador NOT:

```
select * from category
where not catid=1
order by 1;

catid | catgroup |  catname  |                 catdesc
-------+----------+-----------+--------------------------------------------
```

```
2 | Sports   | NHL       | National Hockey League
3 | Sports   | NFL       | National Football League
4 | Sports   | NBA       | National Basketball Association
5 | Sports   | MLS       | Major League Soccer
...
```

En el siguiente ejemplo se usa una condición NOT seguida de una condición AND:

```
select * from category
where (not catid=1) and catgroup='Sports'
order by catid;

catid | catgroup | catname |             catdesc
-------+----------+---------+--------------------------------
2 | Sports   | NHL     | National Hockey League
3 | Sports   | NFL     | National Football League
4 | Sports   | NBA     | National Basketball Association
5 | Sports   | MLS     | Major League Soccer
(4 rows)
```

## Condiciones de coincidencia de patrones

Temas

- LIKE
- SIMILAR TO
- Operadores POSIX

Un operador de coincidencia de patrones busca una cadena para un patrón especificado en la expresión condicional y devuelve true o false según si encuentra una coincidencia o no. Amazon Redshift utiliza tres métodos de coincidencia de patrones:

- Expresiones LIKE

  El operador LIKE compara una expresión de cadena, como el nombre de una columna, con un patrón que usa caracteres comodines % (porcentaje) y _ (guion bajo). La coincidencia de patrones LIKE siempre cubre la cadena completa. LIKE realiza una coincidencia que distingue entre mayúsculas y minúsculas, mientras que ILIKE realiza una coincidencia que no distingue entre mayúsculas y minúsculas.
- Expresiones regulares SIMILAR TO

El operador SIMILAR TO relaciona una expresión de cadena con un patrón de expresión regular de SQL estándar, que puede incluir un conjunto de metacaracteres de coincidencia de patrón que incluyen los dos admitidos por el operador LIKE. SIMILAR TO relaciona la cadena completa y realiza una coincidencia que distingue entre mayúsculas y minúsculas.

- Expresiones regulares de estilo POSIX

  Las expresiones regulares POSIX proporcionan un medio más potente para la coincidencia de patrones que los operadores LIKE y SIMILAR TO. Los patrones de la expresión regular POSIX pueden relacionar cualquier parte de la cadena y realizan una coincidencia que distingue entre mayúsculas y minúsculas.

La coincidencia de expresiones regulares, a través de los operadores SIMILAR TO o POSIX, es costosa informáticamente. Recomendamos utilizar LIKE cuando sea posible, especialmente cuando procesa una gran cantidad de filas. Por ejemplo, las siguientes consultas son idénticas desde el punto de vista funcional, pero la consulta que utiliza LIKE se ejecuta varias veces más rápido que la consulta que utiliza una expresión regular:

```
select count(*) from event where eventname SIMILAR TO '%(Ring|Die)%';
select count(*) from event where eventname LIKE '%Ring%' OR eventname LIKE '%Die%';
```

LIKE

El operador LIKE compara una expresión de cadena, como el nombre de una columna, con un patrón que usa caracteres comodines % (porcentaje) y _ (guion bajo). La coincidencia de patrones LIKE siempre cubre la cadena completa. Para relacionar una secuencia en cualquier lugar dentro de una cadena, el patrón debe comenzar y finalizar con un signo de porcentaje.

LIKE distingue entre mayúsculas y minúsculas; ILIKE no distingue entre mayúsculas y minúsculas.

Sintaxis

```
expression [ NOT ] LIKE | ILIKE pattern [ ESCAPE 'escape_char' ]
```

Argumentos

expresión

Una expresión de carácter UTF-8 válido, como un nombre de columna.

LIKE | ILIKE

LIKE realiza una coincidencia de patrones que distingue entre mayúsculas y minúsculas. ILIKE ejecuta una coincidencia de patrones sin distinción entre mayúsculas y minúsculas para caracteres UTF-8 (ASCII) de un byte. Para ejecutar una coincidencia de patrones sin distinguir entre mayúsculas y minúsculas con caracteres multibyte, utilice la función LOWER de expresión y patrón con una condición LIKE.

Al contrario que los predicados de comparación, como = y <>, los predicados LIKE e ILIKE no omiten implícitamente los espacios finales. Para omitir los espacios finales, utilice RTRIM o convierta explícitamente una columna CHAR en VARCHAR.

El operador ~~ es equivalente a LIKE y ~~* es equivalente a ILIKE. Además, los operadores !~~ y !~~* son equivalentes a NOT LIKE y NOT ILIKE.

pattern

Una expresión de carácter UTF-8 válido con el patrón que se relacionará.

escape_char (carácter_de_escape)

Una expresión de carácter que aplicará escape a metacaracteres en el patrón. La predeterminada es dos barras diagonales invertidas ("\\").

Si el patrón no contiene metacaracteres, solo representa la propia cadena; en ese caso, LIKE actúa igual que el operador de igualdad.

Cualquiera de las expresiones de carácter pueden ser tipos de datos CHAR o VARCHAR. Si son diferentes, Amazon Redshift convierte el patrón al tipo de datos de la expresión.

LIKE admite los siguientes metacaracteres de coincidencia de patrón:

| Operador | Descripción |
|---|---|
| % | Coincide con cualquier secuencia de cero o más caracteres. |
| _ | Coincide con cualquier carácter. |

Ejemplos

En la tabla siguiente se muestran ejemplos de coincidencia de patrones a través de LIKE:

| Expression | Devuelve |
| --- | --- |
| `'abc' LIKE 'abc'` | True |
| `'abc' LIKE 'a%'` | True |
| `'abc' LIKE '_B_'` | False |
| `'abc' ILIKE '_B_'` | True |
| `'abc' LIKE 'c%'` | False |

En el siguiente ejemplo se encuentran todas las ciudades cuyos nombres comienzan con "E":

```
select distinct city from users
where city like 'E%' order by city;
city
---------------
East Hartford
East Lansing
East Rutherford
East St. Louis
Easthampton
Easton
Eatontown
Eau Claire
...
```

En el siguiente ejemplo se encuentran usuarios cuyos apellidos contienen "ten":

```
select distinct lastname from users
where lastname like '%ten%' order by lastname;
lastname
-------------
Christensen
Wooten
...
```

En el siguiente ejemplo se muestra cómo hacer coincidir varios patrones.

```
select distinct lastname from tickit.users
where lastname like 'Chris%' or lastname like '%Wooten' order by lastname;
lastname
-------------
Christensen
Christian
Wooten
...
```

En el siguiente ejemplo se encuentran todas las ciudades cuyos terceros y cuartos caracteres son "ea". El comando usa ILIKE para demostrar que no distingue entre mayúsculas y minúsculas:

```
select distinct city from users where city ilike '__EA%' order by city;
city
-------------
Brea
Clearwater
Great Falls
Ocean City
Olean
Wheaton
(6 rows)
```

En el siguiente ejemplo se usa la cadena de escape predeterminada (\\) para buscar cadenas que incluyan «start_» (el texto `start` seguido de un guion bajo _):

```
select tablename, "column" from pg_table_def
where "column" like '%start\\_%'
limit 5;

     tablename     |    column
-------------------+---------------
 stl_s3client      | start_time
 stl_tr_conflict   | xact_start_ts
 stl_undone        | undo_start_ts
 stl_unload_log    | start_time
 stl_vacuum_detail | start_row
(5 rows)
```

En el siguiente ejemplo se especifica «^» como el carácter de escape y, luego, se utiliza el carácter de escape para buscar cadenas que incluyan «start_» (el texto `start` seguido de un guion bajo _):

```
select tablename, "column" from pg_table_def
where "column" like '%start^_%' escape '^'
limit 5;

     tablename     |    column
-------------------+---------------
 stl_s3client      | start_time
 stl_tr_conflict   | xact_start_ts
 stl_undone        | undo_start_ts
 stl_unload_log    | start_time
 stl_vacuum_detail | start_row
(5 rows)
```

En el siguiente ejemplo, se utiliza el operador ~~* para realizar una búsqueda que no distinga mayúsculas de minúsculas (ILIKE) de ciudades que comiencen por "Ag".

```
select distinct city from users where city ~~* 'Ag%' order by city;

city
------------
Agat
Agawam
Agoura Hills
Aguadilla
```

SIMILAR TO

El operador SIMILAR TO relaciona una expresión de cadena, como el nombre de una columna, con un patrón de expresión regular de SQL estándar. Un patrón de expresión regular de SQL estándar puede incluir un conjunto de metacaracteres de coincidencia de patrón, incluidos los dos admitidos por el operador LIKE.

El operador SIMILAR TO devuelve true solo si su patrón coincide con la cadena completa, a diferencia del comportamiento de la expresión regular POSIX, donde el patrón puede coincidir con cualquier parte de la cadena.

SIMILAR TO realiza una coincidencia de patrones que distingue entre mayúsculas y minúsculas.

> **Note**
>
> La coincidencia de expresiones regulares a través de SIMILAR TO es costosa informáticamente. Recomendamos utilizar LIKE cuando sea posible, especialmente cuando procesa una gran cantidad de filas. Por ejemplo, las siguientes consultas son idénticas desde el punto de vista funcional, pero la consulta que utiliza LIKE se ejecuta varias veces más rápido que la consulta que utiliza una expresión regular:
>
> ```
> select count(*) from event where eventname SIMILAR TO '%(Ring|Die)%';
> select count(*) from event where eventname LIKE '%Ring%' OR eventname LIKE '%Die
> %';
> ```

Sintaxis

```
expression [ NOT ] SIMILAR TO pattern [ ESCAPE 'escape_char' ]
```

Argumentos

expresión

Una expresión de carácter UTF-8 válido, como un nombre de columna.

SIMILAR TO

SIMILAR TO ejecuta una coincidencia de patrones con distinción entre mayúsculas y minúsculas en la cadena completa de expresión.

pattern

Una expresión de carácter UTF-8 válido que representa un patrón de expresión regular de SQL estándar.

escape_char (carácter_de_escape)

Una expresión de carácter que aplicará escape a metacaracteres en el patrón. La predeterminada es dos barras diagonales invertidas ("\\").

Si el patrón no contiene metacaracteres, el patrón solo representa la propia cadena.

Cualquiera de las expresiones de carácter pueden ser tipos de datos CHAR o VARCHAR. Si son diferentes, Amazon Redshift convierte el patrón al tipo de datos de la expresión.

SIMILAR TO admite los siguientes metacaracteres de coincidencia de patrón:

| Operador | Descripción |
|---|---|
| `%` | Coincide con cualquier secuencia de cero o más caracteres. |
| `_` | Coincide con cualquier carácter. |
| `\|` | Denota una alternancia (cualquiera de dos alternativas). |
| `*` | Repite el elemento anterior cero o más veces. |
| `+` | Repite el elemento anterior una o más veces. |
| `?` | Repite el elemento anterior cero o una vez. |
| `{m}` | Repite el elemento anterior exactamente m veces. |
| `{m,}` | Repite el elemento anterior m o más veces. |
| `{m,n}` | Repite el elemento anterior al menos m y no más de n veces. |
| `()` | Los paréntesis agrupan elementos en un único elemento lógico. |
| `[...]` | Una expresión de corchetes especifica una clase de carácter, como en las expresiones regulares POSIX. |

Ejemplos

En la siguiente tabla se muestran ejemplos de coincidencia de patrones a través de SIMILAR TO:

| Expression | Devuelve |
|---|---|
| `'abc' SIMILAR TO 'abc'` | True |
| `'abc' SIMILAR TO '_b_'` | True |
| `'abc' SIMILAR TO '_A_'` | False |
| `'abc' SIMILAR TO '%(b\|d)%'` | True |

| Expression | Devuelve |
| --- | --- |
| `'abc' SIMILAR TO '(b\|c)%'` | False |
| `'AbcAbcdefgefg12efgefg12' SIMILAR TO '((Ab)?c)+d((efg)+(12))+'` | True |
| `'aaaaaab11111xy' SIMILAR TO 'a{6}_[0-9]{5}(x\|y){2}'` | True |
| `'$0.87' SIMILAR TO '$[0-9]+(.[0-9][0-9])?'` | True |

En el siguiente ejemplo, se encuentran las ciudades cuyos nombres contienen “E” o “H”:

```
SELECT DISTINCT city FROM users
WHERE city SIMILAR TO '%E%|%H%' ORDER BY city LIMIT 5;

      city
-----------------
 Agoura Hills
 Auburn Hills
 Benton Harbor
 Beverly Hills
 Chicago Heights
```

En el siguiente ejemplo se usa la cadena de escape predeterminada ("\\") para buscar cadenas que incluyan "_":

```
SELECT tablename, "column" FROM pg_table_def
WHERE "column" SIMILAR TO '%start\\_%'
ORDER BY tablename, "column" LIMIT 5;

        tablename         |       column
--------------------------+---------------------
 stcs_abort_idle          | idle_start_time
 stcs_abort_idle          | txn_start_time
 stcs_analyze_compression | start_time
 stcs_auto_worker_levels  | start_level
 stcs_auto_worker_levels  | start_wlm_occupancy
```

En el siguiente ejemplo se especifica "^" como la cadena de escape y, luego, se utiliza la cadena de escape para buscar cadenas que incluyan "_":

```
SELECT tablename, "column" FROM pg_table_def
WHERE "column" SIMILAR TO '%start^_%' ESCAPE '^'
ORDER BY tablename, "column" LIMIT 5;

        tablename         |       column
--------------------------+---------------------
 stcs_abort_idle          | idle_start_time
 stcs_abort_idle          | txn_start_time
 stcs_analyze_compression | start_time
 stcs_auto_worker_levels  | start_level
 stcs_auto_worker_levels  | start_wlm_occupancy
```

### Operadores POSIX

Una expresión regular POSIX es una secuencia de caracteres que especifica un patrón de coincidencia. Una cadena coincide con una expresión regular si es miembro del conjunto regular descrito por la expresión regular.

Las expresiones regulares POSIX proporcionan un medio más potente para la coincidencia de patrones que los operadores LIKE y SIMILAR TO. Los patrones de expresión regular POSIX pueden coincidir con cualquier parte de una cadena, a diferencia del operador SIMILAR TO, que devuelve true solo si su patrón coincide con la cadena completa.

> **Note**
>
> La coincidencia de expresiones regulares a través de los operadores POSIX es costosa informáticamente. Recomendamos utilizar LIKE cuando sea posible, especialmente cuando procesa una gran cantidad de filas. Por ejemplo, las siguientes consultas son idénticas desde el punto de vista funcional, pero la consulta que utiliza LIKE se ejecuta varias veces más rápido que la consulta que utiliza una expresión regular:
>
> ```
> select count(*) from event where eventname ~ '.*(Ring|Die).*';
> select count(*) from event where eventname LIKE '%Ring%' OR eventname LIKE '%Die
> %';
> ```

Sintaxis

```
expression [ ! ] ~ pattern
```

Argumentos

expresión

Una expresión de carácter UTF-8 válido, como un nombre de columna.

!

Operador de negación. No coincide con la expresión regular.

~

Realiza una coincidencia que distingue entre mayúsculas y minúsculas para cualquier subcadena de expresión.

> **Note**
>
> Un ~~ es sinónimo de LIKE.

pattern

Un literal de cadena que representa un patrón de expresión regular.

Si el patrón no contiene caracteres comodín, el patrón solamente representa la propia cadena.

Para buscar cadenas que incluyan metacaracteres, como `". * | ? "`, etc., aplique escape al carácter con dos barras diagonales invertidas (`" \\"`). A diferencia de `SIMILAR TO` y `LIKE`, la sintaxis de expresión regular POSIX no admite un carácter de escape definido por el usuario.

Cualquiera de las expresiones de carácter pueden ser tipos de datos CHAR o VARCHAR. Si son diferentes, Amazon Redshift convierte el patrón al tipo de datos de la expresión.

Todas las expresiones de carácter pueden ser tipos de datos CHAR o VARCHAR. Si las expresiones tienen diferentes tipos de datos, Amazon Redshift los convierte al tipo de datos de la expresión.

La coincidencia de patrones POSIX admite los siguientes metacaracteres:

| POSIX | Descripción |
| --- | --- |
| . | Coincide con cualquier carácter. |
| * | Coincide con cero o más ocurrencias. |
| + | Coincide con una o más ocurrencias. |
| ? | Coincide con cero o una ocurrencia. |
| \| | Especifica coincidencias alternativas; por ejemplo, E \| H significa E o H. |
| ^ | Coincide con el carácter de comienzo de línea. |
| $ | Coincide con el carácter de final de línea. |
| $ | Coincide con el final de la cadena. |
| [] | Los corchetes especifican una lista de coincidencias, que debe coincidir con una expresión en la lista. Un signo de intercalación (^) precede una lista no coincidente, que relaciona cualquier carácter excepto las expresion es representadas en la lista. |
| ( ) | Los paréntesis agrupan elementos en un único elemento lógico. |
| {m} | Repite el elemento anterior exactamente m veces. |
| {m,} | Repite el elemento anterior m o más veces. |
| {m,n} | Repite el elemento anterior al menos m y no más de n veces. |
| [: :] | Relacione cualquier carácter dentro de una clase de caracteres POSIX. En las siguientes clases de caracteres, Amazon Redshift solo admite caractere s ASCII: [:alnum:] , [:alpha:] , [:lower:] y [:upper:] . |

Amazon Redshift admite las siguientes clases de caracteres POSIX.

| Clase de caracteres | Descripción |
|---|---|
| [[:alnum:]] | Todos los caracteres alfanuméricos ASCII |
| [[:alpha:]] | Todos los caracteres alfabéticos ASCII |
| [[:blank:]] | Todos los caracteres de espacio en blanco |
| [[:cntrl:]] | Todos los caracteres de control (no imprimibles) |
| [[:digit:]] | Todos los dígitos numéricos |
| [[:lower:]] | Todos los caracteres alfabéticos ASCII en minúsculas |
| [[:punct:]] | Todos los caracteres de puntuación |
| [[:space:]] | Todos los caracteres de espacio (no imprimibles) |
| [[:upper:]] | Todos los caracteres alfabéticos ASCII en mayúsculas |
| [[:xdigit:]] | Todos los caracteres hexadecimales válidos |

Amazon Redshift admite los siguientes operadores influenciados por Perl en expresiones regulares. Aplique escape al operador utilizando dos barras diagonales invertidas ("\\’).

| Operador | Descripción | Expresión de clase de carácter equivalente |
|---|---|---|
| \\d | Un carácter que contiene un dígito | [[:digit:]] |
| \\D | Un carácter que no contiene dígitos | [^[:digit:]] |
| \\w | Un carácter que contiene una palabra | [[:word:]] |
| \\W | Un carácter que no contiene palabras | [^[:word:]] |
| \\s | Un carácter de espacio en blanco | [[:space:]] |

| Operador | Descripción | Expresión de clase de carácter equivalente |
|---|---|---|
| \\S | Un carácter que contiene un espacio en blanco | [^[:space:]] |
| \\b | Una palabra límite | |

Ejemplos

En la tabla siguiente se muestran ejemplos de coincidencia de patrones a través de operadores POSIX:

| Expression | Devuelve |
|---|---|
| 'abc' ~ 'abc' | True |
| 'abc' ~ 'a' | True |
| 'abc' ~ 'A' | False |
| 'abc' ~ '.*(b\|d).*' | True |
| 'abc' ~ '(b\|c).*' | True |
| 'AbcAbcdefgefg12efgefg12' ~ '((Ab)?c)+d((efg)+(12))+' | True |
| 'aaaaaab11111xy' ~ 'a{6}.[1]{5}(x\|y){2}' | True |
| '$0.87' ~ '\\$[0-9]+(\\.[0-9][0-9])?' | True |
| 'ab c' ~ '[[:space:]]' | True |
| 'ab c' ~ '\\s' | True |
| ' ' ~ '\\S' | False |

En el siguiente ejemplo, se encuentran las ciudades cuyos nombres contienen E o H:

```
SELECT DISTINCT city FROM users
WHERE city ~ '.*E.*|.*H.*' ORDER BY city LIMIT 5;

      city
-----------------
 Agoura Hills
 Auburn Hills
 Benton Harbor
 Beverly Hills
 Chicago Heights
```

En el siguiente ejemplo, se encuentran las ciudades cuyos nombres no contienen E ni H:

```
SELECT DISTINCT city FROM users WHERE city !~ '.*E.*|.*H.*' ORDER BY city LIMIT 5;

      city
-----------------
 Aberdeen
 Abilene
 Ada
 Agat
 Agawam
```

En el siguiente ejemplo se usa la cadena de escape predeterminada ("\\") para buscar cadenas que incluyan un punto.

```
SELECT venuename FROM venue
WHERE venuename ~ '.*\\..*'
ORDER BY venueid;

          venuename
------------------------------
 St. Pete Times Forum
 Jobing.com Arena
 Hubert H. Humphrey Metrodome
 U.S. Cellular Field
 Superpages.com Center
 E.J. Nutter Center
 Bernard B. Jacobs Theatre
 St. James Theatre
```

## Condición de rango BETWEEN

Una condición BETWEEN prueba expresiones para incluirlas en un rango de valores, con las palabras clave BETWEEN y AND.

### Sintaxis

```
expression [ NOT ] BETWEEN expression AND expression
```

Las expresiones pueden ser tipos de datos de fecha y hora, numéricos o caracteres, pero deben ser compatibles. El rango es inclusivo.

### Ejemplos

El primer ejemplo cuenta cuántas transacciones registraron ventas de 2, 3 o 4 tickets:

```
select count(*) from sales
where qtysold between 2 and 4;

count
--------
104021
(1 row)
```

La condición de rango incluye los valores de inicio y final.

```
select min(dateid), max(dateid) from sales
where dateid between 1900 and 1910;

min  | max
-----+-----
1900 | 1910
```

La primera expresión en una condición de rango debe ser el valor más bajo y la segunda expresión, el valor más alto. En el siguiente ejemplo SIEMPRE se devuelven cero filas debido a los valores de las expresiones:

```
select count(*) from sales
where qtysold between 4 and 2;
```

```
count
-------
0
(1 row)
```

Sin embargo, aplicar el modificador NOT invertirá la lógica y producirá un conteo de todas las filas:

```
select count(*) from sales
where qtysold not between 4 and 2;

count
--------
172456
(1 row)
```

La siguiente consulta devuelve una lista de lugares que tienen entre 20 000 y 50 000 asientos:

```
select venueid, venuename, venueseats from venue
where venueseats between 20000 and 50000
order by venueseats desc;

venueid |       venuename               | venueseats
---------+--------------------------------+------------
116 | Busch Stadium                 |      49660
106 | Rangers BallPark in Arlington |      49115
96 | Oriole Park at Camden Yards   |      48876
...
(22 rows)
```

En el siguiente ejemplo, se demuestra el uso de BETWEEN para valores de fecha:

```
select salesid, qtysold, pricepaid, commission, saletime
from sales
where eventid between 1000 and 2000
   and saletime between '2008-01-01' and '2008-01-03'
order by saletime asc;

salesid | qtysold | pricepaid | commission |   saletime
--------+---------+-----------+------------+---------------
  65082 |       4 |       472 |       70.8 | 1/1/2008 06:06
 110917 |       1 |       337 |      50.55 | 1/1/2008 07:05
```

```
 112103 |       1 |       241 |      36.15 | 1/2/2008 03:15
 137882 |       3 |      1473 |     220.95 | 1/2/2008 05:18
  40331 |       2 |        58 |        8.7 | 1/2/2008 05:57
 110918 |       3 |      1011 |     151.65 | 1/2/2008 07:17
  96274 |       1 |       104 |       15.6 | 1/2/2008 07:18
 150499 |       3 |       135 |      20.25 | 1/2/2008 07:20
  68413 |       2 |       158 |       23.7 | 1/2/2008 08:12
```

Tenga en cuenta que, aunque el intervalo de BETWEEN es inclusivo, las fechas tienen un valor de hora predeterminado de 00:00:00. La única fila válida del 3 de enero para la consulta de ejemplo sería una fila con un valor de saletime de `1/3/2008 00:00:00`.

## Condición nula

La condición nula realiza una prueba en busca de valores nulos, cuando hay un valor que falta o un valor desconocido.

### Sintaxis

```
expression IS [ NOT ] NULL
```

### Argumentos

expresión

Cualquier expresión, como una columna.

IS NULL

Es true cuando el valor de la expresión es nulo y false cuando tiene un valor.

IS NOT NULL

Es false cuando el valor de la expresión es nulo y true cuando tiene un valor.

### Ejemplo

Este ejemplo indica cuántas veces la tabla SALES contiene un valor nulo en el campo QTYSOLD:

```
select count(*) from sales
where qtysold is null;
count
```

```
-------
0
(1 row)
```

## Condición EXISTS

Las condiciones EXISTS realizan pruebas en busca de la existencia de filas en una subconsulta, y devuelve true si una subconsulta devuelve al menos una fila. Si se especifica NOT, la condición devuelve true si una subconsulta no devuelve filas.

### Sintaxis

```
[ NOT ] EXISTS (table_subquery)
```

### Argumentos

EXISTS

Es true cuando table_subquery (subconsulta_de_tabla) devuelve al menos una fila.

NOT EXISTS

Es true cuando table_subquery (subconsulta_de_tabla) no devuelve filas.

table_subquery (subconsulta_de_tabla)

Una subconsulta que toma el valor de una tabla con una o más columnas y una o más filas.

### Ejemplo

Este ejemplo devuelve todos los identificadores de fecha, uno a la vez, para cada fecha que tuvo una venta de cualquier tipo:

```
select dateid from date
where exists (
select 1 from sales
where date.dateid = sales.dateid
)
order by dateid;

dateid
--------
```

```
1827
1828
1829
...
```

## Condición IN

Una condición IN prueba un valor para la membresía en un conjunto de valores o en una subconsulta.

### Sintaxis

```
expression [ NOT ] IN (expr_list | table_subquery)
```

### Argumentos

expresión

Expresión temporal, de carácter o numérica que se compara con expr_list (lista_de_expresiones) o table_subquery (subconsulta_de_tabla) y debe ser compatible con el tipo de datos de esa lista o subconsulta.

expr_list (lista_de_expresiones)

Una o más expresiones separadas por comas o uno o más conjuntos de expresiones separados por comas entre paréntesis.

table_subquery (subconsulta_de_tabla)

Una subconsulta que toma el valor de una tabla con una o más filas, pero está limitada a una columna en su lista selecta.

IN | NOT IN

IN devuelve true si la expresión es un miembro de la consulta o lista de expresiones. NOT IN devuelve true si la expresión no es un miembro. IN y NOT IN devuelven NULL y no devuelven filas en los siguientes casos: si la expresión genera un valor nulo o si no hay valores de expr_list o table_subquery que coincidan y al menos una de estas filas de comparación genera un valor nulo.

### Ejemplos

Las siguientes condiciones son true solo para esos valores enumerados:

```
qtysold in (2, 4, 5)
date.day in ('Mon', 'Tues')
date.month not in ('Oct', 'Nov', 'Dec')
```

Optimización para listas IN grandes

Para optimizar el rendimiento de la consulta, una lista IN que incluye más de 10 valores se evalúa internamente como una matriz escalar. Las listas IN con menos de 10 valores se evalúan como una serie de predicados OR. Esta optimización se admite para los tipos de datos SMALLINT, INTEGER, BIGINT, REAL, DOUBLE PRECISION, BOOLEAN, CHAR, VARCHAR, DATE, TIMESTAMP y TIMESTAMPTZ.

Observe el resultado de EXPLAIN de la consulta para ver el efecto de esta optimización. Por ejemplo:

```
explain select * from sales
QUERY PLAN
------------------------------------------------------------------------
XN Seq Scan on sales  (cost=0.00..6035.96 rows=86228 width=53)
Filter: (salesid = ANY ('{1,2,3,4,5,6,7,8,9,10,11}'::integer[]))
(2 rows)
```

# Comandos SQL

El lenguaje SQL consta de comandos que se utilizan para crear y manipular objetos de base de datos, ejecutar consultas, cargar tablas y modificar los datos de las tablas.

Amazon Redshift se basa en PostgreSQL. Amazon Redshift y PostgreSQL tienen una serie de diferencias significativas que debe tener en cuenta a la hora de diseñar y desarrollar aplicaciones de almacenamiento de datos. Para obtener más información acerca de las diferencias entre Amazon Redshift SQL y PostgreSQL, consulte Amazon Redshift y PostgreSQL.

> **Note**
>
> El tamaño máximo de una instrucción SQL es de 16 MB.

Temas

- ABORT
- ALTER DATABASE
- ALTER DATASHARE
- ALTER DEFAULT PRIVILEGES
- ALTER EXTERNAL SCHEMA
- ALTER EXTERNAL VIEW
- ALTER FUNCTION
- ALTER GROUP
- MODIFICAR PROVEEDOR DE IDENTIDADES
- ALTER MASKING POLICY
- ALTER MATERIALIZED VIEW
- ALTER RLS POLICY
- ALTER ROLE
- ALTER PROCEDURE
- ALTER SCHEMA
- ALTER SYSTEM
- ALTER TABLE
- ALTER TABLE APPEND
- ALTER USER
- ANALYZE
- ANALYZE COMPRESSION
- ATTACH MASKING POLICY
- ATTACH RLS POLICY
- BEGIN
- CALL
- CANCEL
- CLOSE
- COMMENT

- COMMIT
- COPY
- CREATE DATABASE
- CREATE DATASHARE
- CREATE EXTERNAL FUNCTION
- CREATE EXTERNAL MODEL
- CREATE EXTERNAL SCHEMA
- CREATE EXTERNAL TABLE
- CREATE EXTERNAL VIEW
- CREATE FUNCTION
- CREATE GROUP
- CREATE IDENTITY PROVIDER
- CREATE LIBRARY
- CREATE MASKING POLICY
- CREATE MATERIALIZED VIEW
- CREATE MODEL
- CREATE PROCEDURE
- CREATE RLS POLICY
- CREAR ROL
- CREATE SCHEMA
- CREATE TABLE
- CREATE TABLE AS
- CREAR USUARIO
- CREATE VIEW
- DEALLOCATE
- DECLARE
- DELETE
- DESC DATASHARE

- DESC IDENTITY PROVIDER
- DETACH MASKING POLICY
- DETACH RLS POLICY
- DROP DATABASE
- DROP DATASHARE
- DROP EXTERNAL VIEW
- DROP FUNCTION
- DROP GROUP
- DROP IDENTITY PROVIDER
- DROP LIBRARY
- DROP MASKING POLICY
- DROP MODEL
- DROP MATERIALIZED VIEW
- DROP PROCEDURE
- DROP RLS POLICY
- DROP ROLE
- DROP SCHEMA
- DROP TABLE
- DROP USER
- DROP VIEW
- END
- EXECUTE
- EXPLAIN
- FETCH
- GRANT
- INSERT
- INSERT (tabla externa)
- LOCK

- MERGE
- PREPARE
- REFRESH MATERIALIZED VIEW
- RESET
- REVOKE
- ROLLBACK
- SELECT
- SELECT INTO
- SET
- SET SESSION AUTHORIZATION
- SET SESSION CHARACTERISTICS
- SHOW
- SHOW COLUMNS
- SHOW EXTERNAL TABLE
- SHOW DATABASES
- SHOW GRANTS
- SHOW MODEL
- SHOW DATASHARES
- SHOW PROCEDURE
- MOSTRAR ESQUEMAS
- SHOW TABLE
- SHOW TABLES
- SHOW VIEW
- START TRANSACTION
- TRUNCATE
- UNLOAD
- UPDATE
- USE

- VACUUM

# ABORT

Detiene la transacción que se está ejecutando en ese momento y descarta todas las actualizaciones realizadas por esa transacción. ABORT no tiene efecto en transacciones que ya están completadas.

Este comando lleva a cabo la misma función que el comando ROLLBACK. Para obtener más información, consulta ROLLBACK.

## Sintaxis

```
ABORT [ WORK | TRANSACTION ]
```

## Parámetros

WORK

Palabra clave opcional.

TRANSACTION

Palabra clave opcional; WORK y TRANSACTION son sinónimos.

## Ejemplo

En el siguiente ejemplo, se crea una tabla que inicia una transacción donde los datos se insertan en la tabla. Luego, el comando ABORT revierte la inserción de datos para dejar la tabla vacía.

El siguiente comando crea una tabla de ejemplo denominada MOVIE_GROSS:

```
create table movie_gross( name varchar(30), gross bigint );
```

El siguiente conjunto de comandos inicia una transacción en donde se insertan dos filas de datos en la tabla:

```
begin;

insert into movie_gross values ( 'Raiders of the Lost Ark', 23400000);
```

```
insert into movie_gross values ( 'Star Wars', 10000000 );
```

A continuación, el siguiente comando selecciona los datos de la tabla para mostrar que se insertaron correctamente:

```
select * from movie_gross;
```

El resultado del comando muestra que ambas filas se insertaron correctamente:

```
         name           |  gross
------------------------+----------
Raiders of the Lost Ark | 23400000
Star Wars               | 10000000
(2 rows)
```

Este comando ahora revierte los cambios de datos para que vuelvan al estado de inicio de la transacción:

```
abort;
```

Si selecciona los datos de la tabla, ahora se muestra una tabla vacía:

```
select * from movie_gross;

 name | gross
------+-------
(0 rows)
```

# ALTER DATABASE

Cambia los atributos de una base de datos.

## Privilegios necesarios

Para utilizar ALTER DATABASE, se requiere uno de los siguientes privilegios.

- Superusuario
- Usuarios con el privilegio ALTER DATABASE
- Propietario de la base de datos

## Sintaxis

```
ALTER DATABASE database_name
{
  RENAME TO new_name
  | OWNER TO new_owner
  | [ CONNECTION LIMIT { limit | UNLIMITED } ]
    [ COLLATE { CASE_SENSITIVE | CASE_INSENSITIVE } ]
    [ ISOLATION LEVEL { SERIALIZABLE | SNAPSHOT } ]
| INTEGRATION
 {
  REFRESH { { ALL | INERROR } TABLES [ IN SCHEMA schema [, ...] ] | TABLE schema.table
 [, ...] }
   | SET
     [ QUERY_ALL_STATES [=] { TRUE | FALSE } ]
     [ ACCEPTINVCHARS [=] { TRUE | FALSE } ]
     [ REFRESH_INTERVAL <interval> ]
     [ TRUNCATECOLUMNS [=] { TRUE | FALSE } ]
     [ HISTORY_MODE [=] {TRUE | FALSE} [ FOR { {ALL} TABLES [IN SCHEMA schema [, ...] ]
 | TABLE schema.table [, ...] } ] ]
 }
}
```

## Parámetros

database_name

Nombre de la base de datos que se modificará. Por lo general, modifica una base de datos a la que no está actualmente conectado. De todos modos, los cambios entran en vigor únicamente en sesiones posteriores. Puede cambiar el propietario de la base de datos actual, pero no puede cambiar su nombre:

```
alter database tickit rename to newtickit;
ERROR:  current database may not be renamed
```

RENAME TO

Cambia el nombre de la base de datos especificada. Para obtener más información acerca de los nombres válidos, consulte Nombres e identificadores. No puede cambiar el nombre de las bases de datos dev, padb_harvest, template0, template1 o sys:internal, ni puede cambiar el nombre de la base de datos actual. Solo el propietario de la base de datos o un superuser (p. 984) puede

cambiar el nombre de una base de datos; los propietarios que no son superusuarios deben tener también el privilegio CREATEDB.

new_name

Nuevo nombre de la base de datos.

OWNER TO

Cambia el propietario de la base de datos especificada. Puede cambiar el propietario de la base de datos actual o alguna otra base de datos. Solo un superusuario puede cambiar el propietario.

new_owner

Nuevo propietario de la base de datos. El nuevo propietario debe ser un usuario de la base de datos existente con privilegios de escritura. Para obtener más información acerca de los privilegios del usuario, consulte GRANT.

CONNECTION LIMIT { limit | UNLIMITED }

La cantidad máxima de conexiones a la base de datos que los usuarios pueden tener abiertas al mismo tiempo. Este límite no se aplica a los superusuarios. Use la palabra clave UNLIMITED para permitir la cantidad máxima de conexiones simultáneas. También puede aplicarse un límite de la cantidad de conexiones de cada usuario. Para obtener más información, consulte CREAR USUARIO. El valor predeterminado es UNLIMITED. Para ver las conexiones actuales, consulte la vista del sistema STV_SESSIONS.

> **Note**
>
> Si se aplican los límites de conexión tanto para usuarios como para bases de datos, debe haber una ranura de conexión sin utilizar disponible dentro de ambos límites cuando un usuario intenta conectarse.

COLLATE { CASE_SENSITIVE | CASE_INSENSITIVE }

Se trata de una cláusula que especifica si la búsqueda o la comparación de cadenas distinguen o no entre letras mayúsculas y minúsculas.

Puede modificar la distinción entre mayúsculas y minúsculas de la base de datos actual, que está vacía.

Debe contar con el privilegio sobre la base de datos actual para cambiar la distinción entre letras mayúsculas y minúsculas. Los superusuarios o propietarios de bases de datos con el privilegio

CREATE DATABASE también pueden cambiar la capacidad de distinguir entre mayúsculas y minúsculas de las base de datos.

ISOLATION LEVEL { SERIALIZABLE | SNAPSHOT }

Una cláusula que especifica el nivel de aislamiento utilizado cuando las consultas se ejecutan en una base de datos.

- Aislamiento SERIALIZABLE: proporciona serialización completa para transacciones simultáneas. Para obtener más información, consulte Aislamiento serializable.
- Aislamiento SNAPSHOT: proporciona un nivel de aislamiento con protección contra conflictos de actualización y eliminación.

Para obtener más información acerca de los niveles de aislamiento, consulte CREATE DATABASE.

Tenga en cuenta los siguientes aspectos al modificar el nivel de aislamiento de una base de datos:

- Debe contar con el privilegio de superusuario o CREATE DATABASE en la base de datos actual para cambiar el nivel de aislamiento de la base de datos.
- No puede modificar el nivel de aislamiento de la base de datos de `dev`.
- No puede modificar el nivel de aislamiento dentro de un bloque de transacción.
- El comando alter isolation level produce un error si otros usuarios están conectados a la base de datos.
- El comando alter isolation level puede modificar la configuración del nivel de aislamiento de la sesión actual.

INTEGRATION

Modifique una base de datos de integración sin ETL.

REFRESH {{ ALL | INERROR } TABLES [IN SCHEMA esquema [, ...]] | TABLE esquema.tabla [, ...]}

Una cláusula que especifica si Amazon Redshift actualizará todas las tablas o las que contengan errores en el esquema o la tabla especificados. La actualización hará que las tablas del esquema o tabla especificados se repliquen completamente desde la base de datos de origen.

Para obtener más información, consulte Integraciones sin ETL en la Guía de administración de Amazon Redshift. Para obtener más información acerca de los estados de integración, consulte SVV_INTEGRATION_TABLE_STATE y SVV_INTEGRATION.

QUERY_ALL_STATES [=] { TRUE | FALSE }

La cláusula QUERY_ALL_STATES establece si las tablas de integración sin ETL pueden consultarse en todos los estados (`Synced`, `Failed`, `ResyncRequired` y `ResyncInitiated`). De forma predeterminada, una tabla de integración sin ETL solo puede consultarse en estado `Synced`.

ACCEPTINVCHARS [=] { TRUE | FALSE }

La cláusula ACCEPTINVCHARS establece si las tablas de integración sin ETL continúan con la ingesta cuando se detectan caracteres no válidos para el tipo de datos VARCHAR. Cuando se encuentran caracteres no válidos, se reemplazan por un carácter ? predeterminado.

REFRESH_INTERVAL <intervalo>

La cláusula REFRESH_INTERVAL establece el intervalo de tiempo aproximado, en segundos, para actualizar los datos desde el origen sin ETL en la base de datos de destino. El `interval` se puede establecer de 0 a 432 000 segundos (5 días) para las integraciones sin ETL cuyo tipo de origen sea Aurora MySQL, Aurora PostgreSQL o RDS para MySQL. Para las integraciones sin ETL de Amazon DynamoDB, se puede configurar el `interval` entre 900 y 432 000 segundos (15 minutos y 5 días).

Para obtener más información sobre la creación de bases de datos con integraciones sin ETL, consulte Creación de bases de datos de destino en Amazon Redshift en la Guía de administración de Amazon Redshift.

TRUNCATECOLUMNS [=] { TRUE | FALSE }

La cláusula TRUNCATECOLUMNS establece si las tablas de integración sin ETL continúan con la ingesta cuando los valores de la columna VARCHAR o los atributos de columna SUPER exceden el límite. Cuando `TRUE`, los valores se truncan para que quepan en la columna y los valores de los atributos JSON desbordados se truncan para que quepan en la columna SUPER.

HISTORY_MODE [=] {TRUE | FALSE} [ FOR { {ALL} TABLES [IN SCHEMA esquema [, ...]] | TABLE esquema.tabla [, ...]} ]

Cláusula que especifica si Amazon Redshift establecerá el modo de historial para todas las tablas o las tablas del esquema especificado que participan en la integración sin ETL. Esta opción solo se aplica a las bases de datos creadas para la integración sin ETL.

La cláusula HISTORY_MODE puede establecerse en `TRUE` o `FALSE`. El valor predeterminado es `FALSE`. Activar y desactivar el modo de historial solo es aplicable a las tablas que se encuentran

en el estado `Synced`. Para obtener información sobre HISTORY_MODE, consulte Modo de historial en la Guía de administración de Amazon Redshift.

## Notas de uso

Los comandos ALTER DATABASE se aplican en sesiones posteriores, no en sesiones actuales. Debe volver a conectarse a la base de datos modificada para ver el efecto del cambio.

## Ejemplos

En el siguiente ejemplo, se cambia el nombre de una base de datos denominada TICKIT_SANDBOX a TICKIT_TEST:

```
alter database tickit_sandbox rename to tickit_test;
```

En el siguiente ejemplo, se cambia el propietario de la base de datos TICKIT (la base de datos actual) a DWUSER:

```
alter database tickit owner to dwuser;
```

En el siguiente ejemplo, se cambia la distinción entre letras mayúsculas y minúsculas de la base de datos sampledb:

```
ALTER DATABASE sampledb COLLATE CASE_INSENSITIVE;
```

En el siguiente ejemplo se modifica una base de datos denominada **`sampledb`** con nivel de aislamiento SNAPSHOT.

```
ALTER DATABASE sampledb ISOLATION LEVEL SNAPSHOT;
```

En el siguiente ejemplo se actualizan las tablas **`schema1.sample_table1`** y **`schema2.sample_table2`** de la base de datos **`sample_integration_db`** de la integración sin ETL.

```
ALTER DATABASE sample_integration_db INTEGRATION REFRESH TABLE schema1.sample_table1,
 schema2.sample_table2;
```

En el siguiente ejemplo se actualizan todas las tablas sincronizadas y con errores de la integración sin ETL.

```
ALTER DATABASE sample_integration_db INTEGRATION REFRESH ALL tables;
```

El siguiente ejemplo establece el intervalo de actualización para las integraciones sin ETL en 600 segundos.

```
ALTER DATABASE sample_integration_db INTEGRATION SET REFRESH_INTERVAL 600;
```

En el siguiente ejemplo se actualizan todas las tablas incluidas en el `ErrorState` del esquema **`sample_schema`**.

```
ALTER DATABASE sample_integration_db INTEGRATION REFRESH INERROR TABLES in SCHEMA
 sample_schema;
```

En el siguiente ejemplo se activa el modo de historial para la tabla `myschema.table1`.

```
ALTER DATABASE sample_integration_db INTEGRATION SET HISTORY_MODE = true FOR TABLE
 myschema.table1
```

En el siguiente ejemplo se activa el modo de historial para todas las tablas en `myschema`.

```
ALTER DATABASE sample_integration_db INTEGRATION SET HISTORY_MODE = true for ALL TABLES
 IN SCHEMA myschema
```

# ALTER DATASHARE

Cambia la definición de un datashare (recurso para compartir datos). Puede agregar objetos o quitarlos con ALTER DATASHARE. Solo puede cambiar un recurso compartido de datos en la base de datos actual. Agregue o elimine objetos de la base de datos asociada a un recurso compartido de datos. El propietario del datashare con los permisos necesarios sobre los objetos del datashare que se agregarán o se quitarán puede modificar el datashare.

## Privilegios necesarios

Los siguientes privilegios son necesarios para ALTER DATASHARE:

- Superusuario.
- Usuario con el privilegio ALTER DATASHARE.
- Usuarios que tienen el privilegio ALTER u ALL en el recurso compartido de datos.

- Para agregar objetos específicos a un recurso compartido de datos, los usuarios deben disponer del privilegio correspondiente sobre los objetos. En este caso, los usuarios deben ser propietarios de los objetos, o bien tener los privilegios SELECT, USAGE u ALL sobre ellos.

## Sintaxis

La sintaxis siguiente ilustra cómo agregar o quitar objetos al recurso compartido de datos.

```
ALTER DATASHARE datashare_name { ADD | REMOVE } {
TABLE schema.table [, ...]
| SCHEMA schema [, ...]
| FUNCTION schema.sql_udf (argtype,...) [, ...]
| ALL TABLES IN SCHEMA schema [, ...]
| ALL FUNCTIONS IN SCHEMA schema [, ...] }
```

La sintaxis siguiente ilustra cómo configurar las propiedades del recurso compartido de datos.

```
ALTER DATASHARE datashare_name {
[ SET PUBLICACCESSIBLE [=] TRUE | FALSE ]
[ SET INCLUDENEW [=] TRUE | FALSE FOR SCHEMA schema ] }
```

## Parámetros

datashare_name

Se trata del nombre del datashare que se modificará.

ADD | REMOVE

Se trata de una cláusula que especifica si se agregan o se quitan objetos del datashare.

TABLE schema.table [, ...]

Se trata del nombre de la tabla o la vista del esquema especificado que se agrega al datashare.

SCHEMA schema [, ...]

Se trata del nombre del esquema que se agrega al datashare.

FUNCTION schema.sql_udf (argtype,...) [, ...]

Se trata del nombre de la función definida por el usuario de SQL con tipos de argumentos que se agrega al recurso compartido de datos.

ALL TABLES IN SCHEMA schema [, ...]

Se trata de una cláusula que especifica si se agregan todas las tablas y las vistas del esquema especificado al datashare.

ALL FUNCTIONS IN SCHEMA schema [, ...] }

Se trata de una cláusula que especifica la incorporación de todas las funciones del esquema especificado al datashare.

[ SET PUBLICACCESSIBLE [=] TRUE | FALSE ]

Se trata de una cláusula que especifica si un datashare se puede compartir con clústeres que sean accesibles públicamente.

[ SET INCLUDENEW [=] TRUE | FALSE FOR SCHEMA schema ]

Se trata de una cláusula que especifica si se agregan futuras tablas, vistas o funciones definidas por el usuario (UDF) de SQL creadas en el esquema especificado al datashare. Las tablas, las vistas o las UDF de SQL actuales del esquema especificado no se agregan al datashare. Solo los superusuarios pueden modificar esta propiedad para cada par de esquema y datashare. De manera predeterminada, la cláusula INCLUDENEW es false.

## Notas de uso de ALTER DATASHARE

- Los siguientes usuarios pueden modificar un datashare:
  - un superusuario
  - el propietario del datashare
  - usuarios que tienen privilegios ALTER o ALL en el recurso compartido de datos
- Para agregar objetos específicos a un recurso compartido de datos, los usuarios deben disponer de los privilegios correspondientes sobre los objetos. Los usuarios deben ser los propietarios de los objetos o tener privilegios SELECT, USAGE u ALL sobre los objetos.
- Puede compartir esquemas, tablas, vistas normales, vistas de enlace de tiempo de ejecución, vistas materializadas y funciones definidas por el usuario (UDF) de SQL. Primero, agregue un esquema al recurso compartido de datos antes de agregar otros objetos al esquema.

  Cuando se agrega un esquema, Amazon Redshift no agrega todos los objetos que contiene. Se deben agregar de forma explícita.
- Le recomendamos que cree recursos compartidos de datos de AWS Data Exchange con la configuración de acceso público activada.

- En general, recomendamos que no modifique un recurso compartido de datos de AWS Data Exchange para desactivar la accesibilidad pública mediante la instrucción ALTER DATASHARE. En caso contrario, las Cuentas de AWS que tengan acceso al recurso compartido de datos lo pierden si sus clústeres son accesibles de manera pública. Realizar este tipo de alteración puede infringir los términos del producto de datos en AWS Data Exchange. Para obtener una excepción a esta recomendación, consulte lo siguiente.

  El siguiente ejemplo muestra un error cuando se crea un recurso compartido de datos de AWS Data Exchange con la configuración desactivada.

```
ALTER DATASHARE salesshare SET PUBLICACCESSIBLE FALSE;
ERROR:  Alter of ADX-managed datashare salesshare requires session variable
 datashare_break_glass_session_var to be set to value 'c670ba4db22f4b'
```

  Para permitir la alteración de un recurso compartido de datos de AWS Data Exchange para desactivar la configuración de acceso público, configure la siguiente variable y ejecute de nuevo la instrucción ALTER DATASHARE.

```
SET datashare_break_glass_session_var to 'c670ba4db22f4b';
```

```
ALTER DATASHARE salesshare SET PUBLICACCESSIBLE FALSE;
```

  En este caso, Amazon Redshift genera un valor aleatorio único para configurar la variable de sesión para permitir ALTER DATASHARE SET PUBLICACCESIBLE FALSE para un recurso compartido de datos de AWS Data Exchange.

## Ejemplos

En el siguiente ejemplo, se agrega el esquema `public` al recurso compartido de datos de `salesshare`.

```
ALTER DATASHARE salesshare ADD SCHEMA public;
```

En el siguiente ejemplo, se agregan todas las tablas de `public.tickit_sales_redshift` al recurso compartido de datos de `salesshare`.

```
ALTER DATASHARE salesshare ADD TABLE public.tickit_sales_redshift;
```

En el siguiente ejemplo, se agregan todas las tablas al recurso compartido de datos de `salesshare`.

```
ALTER DATASHARE salesshare ADD ALL TABLES IN SCHEMA PUBLIC;
```

En el siguiente ejemplo, se elimina la tabla de `public.tickit_sales_redshift` del recurso compartido de datos de `salesshare`.

```
ALTER DATASHARE salesshare REMOVE TABLE public.tickit_sales_redshift;
```

# ALTER DEFAULT PRIVILEGES

Define el conjunto de permisos de acceso predeterminado que se van a aplicar a los objetos que el usuario especificado cree en el futuro. De manera predeterminada, los usuarios pueden cambiar solo sus propios permisos de acceso predeterminados. Solo un superusuario puede especificar los permisos predeterminados de otros usuarios.

Puede aplicar privilegios predeterminados a roles, usuarios o grupos de usuarios. Puede configurar permisos predeterminados globalmente para todos los objetos creados en la base de datos actual o para objetos creados solo en los esquemas especificados.

Los permisos predeterminados se aplican solo a los objetos nuevos. Ejecutar ALTER DEFAULT PRIVILEGES no cambia los permisos de objetos existentes. Para conceder permisos a todos los objetos actuales y futuros creados por cualquier usuario dentro de una base de datos o un esquema, consulte Permisos acotados.

Para ver información acerca de los privilegios predeterminados para usuarios de bases de datos, consulte la tabla de catálogo del sistema PG_DEFAULT_ACL.

Para obtener más información acerca de los privilegios, consulte GRANT.

## Privilegios necesarios

Los siguientes privilegios son necesarios para ALTER DEFAULT PRIVILEGES:

- Superusuario
- Usuarios con el privilegio ALTER DEFAULT PRIVILEGES
- Usuarios que cambien sus propios privilegios de acceso predeterminados

- Usuarios que establezcan privilegios para esquemas para los que tienen privilegios de acceso

## Sintaxis

```
ALTER DEFAULT PRIVILEGES
    [ FOR USER target_user [, ...] ]
    [ IN SCHEMA schema_name [, ...] ]
    grant_or_revoke_clause

where grant_or_revoke_clause is one of:

GRANT { { SELECT | INSERT | UPDATE | DELETE | DROP | REFERENCES | TRUNCATE } [,...] |
 ALL [ PRIVILEGES ] }
 ON TABLES
 TO { user_name [ WITH GRANT OPTION ] | ROLE role_name | GROUP group_name | PUBLIC }
 [, ...]

GRANT { EXECUTE | ALL [ PRIVILEGES ] }
 ON FUNCTIONS
 TO { user_name [ WITH GRANT OPTION ] |  ROLE role_name | GROUP group_name | PUBLIC }
 [, ...]

GRANT { EXECUTE | ALL [ PRIVILEGES ] }
 ON PROCEDURES
 TO { user_name [ WITH GRANT OPTION ] |  ROLE role_name | GROUP group_name | PUBLIC }
 [, ...]

REVOKE [ GRANT OPTION FOR ] { { SELECT | INSERT | UPDATE | DELETE | REFERENCES |
 TRUNCATE } [,...] | ALL [ PRIVILEGES ] }
 ON TABLES
 FROM user_name [, ...] [ RESTRICT ]

REVOKE  { { SELECT | INSERT | UPDATE | DELETE | REFERENCES | TRUNCATE } [,...] | ALL
 [ PRIVILEGES ] }
 ON TABLES
 FROM { ROLE role_name | GROUP group_name | PUBLIC } [, ...] [ RESTRICT ]

REVOKE [ GRANT OPTION FOR ] { EXECUTE | ALL [ PRIVILEGES ] }
 ON FUNCTIONS
 FROM user_name [, ...] [ RESTRICT ]

REVOKE { EXECUTE | ALL [ PRIVILEGES ] }
 ON FUNCTIONS
```

```
 FROM { ROLE role_name | GROUP group_name | PUBLIC } [, ...] [ RESTRICT ]

REVOKE [ GRANT OPTION FOR ] { EXECUTE | ALL [ PRIVILEGES ] }
 ON PROCEDURES
 FROM user_name [, ...] [ RESTRICT ]

REVOKE { EXECUTE | ALL [ PRIVILEGES ] }
 ON PROCEDURES
 FROM { ROLE role_name | GROUP group_name | PUBLIC } [, ...] [ RESTRICT ]
```

## Parámetros

FOR USER target_user

Opcional. El nombre del usuario para el que se definen los privilegios predeterminados. Solo un superusuario puede especificar los privilegios predeterminados de otros usuarios. El valor predeterminado es el usuario actual.

IN SCHEMA schema_name

Opcional. Si aparece una cláusula IN SCHEMA, los privilegios predeterminados especificados se aplican a objetos nuevos creados en el schema_name (nombre_de_esquema) especificado. En este caso, el usuario o grupo de usuarios que es destino de ALTER DEFAULT PRIVILEGES debe tener el privilegio CREATE para el esquema especificado. Los privilegios predeterminados que son específicos de un esquema se agregan a los privilegios predeterminados globales existentes. Por defecto, los privilegios predeterminados se aplican globalmente a toda la base de datos.

GRANT

Define el conjunto de privilegios que se concederá a los usuarios o grupos especificados para todas las tablas y vistas, funciones o procedimientos almacenados nuevos creados por el usuario especificado. Puede configurar los mismos privilegios y opciones con la cláusula GRANT que con el comando GRANT.

WITH GRANT OPTION

Una cláusula que indica que el usuario que recibe los privilegios puede, a cambio, concederles los mismos privilegios a otros. No puede conceder WITH GRANT OPTION a un grupo o a PUBLIC.

TO user_name | ROLE role_name | GROUP group_name

El nombre del usuario, rol o grupo de usuarios a los que se aplican los privilegios predeterminados especificados.

REVOKE

El conjunto de privilegios que se revocará de los usuarios o grupos especificados para todas las tablas, funciones o procedimientos almacenados creados por el usuario especificado. Puede configurar los mismos privilegios y opciones con la cláusula REVOKE que con el comando REVOKE.

GRANT OPTION FOR

Una cláusula que revoca solo la opción de conceder un privilegio especificado a otros usuarios y no revoca el privilegio en sí. No puede revocar GRANT OPTION de un grupo o de PUBLIC.

FROM user_name | ROLE role_name | GROUP group_name

El nombre del usuario o grupo de usuarios de los que se revocan los privilegios especificados de manera predeterminada.

RESTRICT

La opción RESTRICT revoca solo aquellos privilegios que el usuario haya concedido directamente. Esta es la opción predeterminada.

## Ejemplos

Supongamos que desea permitir que cualquier usuario del grupo de usuarios `report_readers` vea todas las tablas y vistas creadas por el usuario `report_admin`. En este caso, ejecute el siguiente comando como superusuario.

```
alter default privileges for user report_admin grant select on tables to group
 report_readers;
```

En el siguiente ejemplo el primer comando concede privilegios SELECT en todas las tablas nuevas que usted crea. Cada vez que cree una vista nueva, debe conceder privilegios a la vista de forma explícita o volver a ejecutar el comando `alter default privileges`.

```
alter default privileges grant select on tables to public;
```

En el siguiente ejemplo, se concede un privilegio INSERT al grupo de usuarios `sales_admin` para todas las tablas y vistas nuevas que usted crea en el esquema `sales`.

```
alter default privileges in schema sales grant insert on tables to group sales_admin;
```

En el siguiente ejemplo, se revierte el comando ALTER DEFAULT PRIVILEGES del ejemplo anterior.

```
alter default privileges in schema sales revoke insert on tables from group
 sales_admin;
```

De manera predeterminada, el grupo de usuarios PUBLIC tiene permiso de ejecución para todas las funciones nuevas definidas por el usuario. Para revocar los permisos de ejecución `public` para las funciones nuevas y, luego, conceder el permiso de ejecución solo al grupo de usuarios `dev_test`, ejecute los siguientes comandos.

```
alter default privileges revoke execute on functions from public;
alter default privileges grant execute on functions to group dev_test;
```

## ALTER EXTERNAL SCHEMA

Modifica un esquema externo existente en la base de datos actual. Solo los propietarios del esquema, los superusuarios o los usuarios con privilegios ALTER en el esquema pueden modificarlo. Solo se pueden modificar los esquemas externos creados a partir de DATA CATALOG, KAFKA o MSK.

El propietario de este esquema es el emisor del comando CREATE EXTERNAL SCHEMA. Para transferir la propiedad de un esquema externo, use ALTER SCHEMA para cambiar el propietario. Para conceder acceso al esquema a otros usuarios o grupos de usuarios, utilice el comando GRANT.

No puede usar los comandos GRANT o REVOKE para los permisos en una tabla externa. En lugar de ello, conceda o revoque los permisos en el esquema externo.

Para obtener más información, consulte los siguientes temas:

- ALTER SCHEMA
- GRANT
- REVOKE
- CREATE EXTERNAL SCHEMA
- Habilitación de la autenticación mTLS para un esquema externo existente

Para ver detalles de los esquemas externos, consulte la vista del sistema SVV_EXTERNAL_SCHEMAS. Para obtener más información, consulte SVV_EXTERNAL_SCHEMAS.

## Sintaxis

```
ALTER EXTERNAL SCHEMA schema_name
[ IAM_ROLE [ default | 'SESSION' | 'arn:aws:iam::<AWS account-id>:role/<role-name>' ] ]
[ AUTHENTICATION [ none | iam | mtls] ]
[ AUTHENTICATION_ARN 'acm-certificate-arn' | SECRET_ARN 'asm-secret-arn' ]
[ URI 'Kafka bootstrap URL' ]
```

Si tiene un esquema externo existente que utiliza para la ingesta de streaming y desea implementar un TLS mutuo para la autenticación, puede ejecutar un comando como el siguiente, que especifica la autenticación mTLS y el ARN del certificado ACM en ACM.

```
ALTER EXTERNAL SCHEMA schema_name
AUTHENTICATION mtls
AUTHENTICATION_ARN 'arn:aws:acm:us-east-1:444455556666:certificate/certificate_ID';
```

O bien, puede modificar la autenticación mTLS, con referencia al ARN del secreto en Secrets Manager.

```
ALTER EXTERNAL SCHEMA schema_name
AUTHENTICATION mtls
SECRET_ARN 'arn:aws:secretsmanager:us-east-1:012345678910:secret:myMTLSSecret';
```

El siguiente ejemplo muestra cómo modificar el URI para ALTER EXTERNAL SCHEMA:

```
ALTER EXTERNAL SCHEMA schema_name
URI 'lkc-ghidef-67890.centralus.azure.glb.confluent.cloud:9092';
```

El siguiente ejemplo muestra cómo modificar el rol de IAM para ALTER EXTERNAL SCHEMA:

```
ALTER EXTERNAL SCHEMA schema_name
IAM_ROLE 'arn:aws:iam::012345678901:role/testrole';
```

## Parámetros

IAM_ROLE[ default | 'SESSION' | 'arn:aws:iam::<AWS account-id>:role/<role-name>' ]

Utilice la palabra clave `default` para que Amazon Redshift utilice el rol de IAM establecido como predeterminado.

Use `'SESSION'` si se conecta al clúster de Amazon Redshift mediante una identidad federada y acceda a las tablas desde el esquema externo creado con este comando.

Para obtener más información, consulte CREATE EXTERNAL SCHEMA.

AUTENTICACIÓN

El tipo de autenticación que se ha definido para la ingesta de streaming. La ingesta de streaming con tipos de autenticación funciona con Apache Kafka, Confluent Cloud y Amazon Managed Streaming para Apache Kafka. Para obtener más información, consulte CREATE EXTERNAL SCHEMA.

AUTHENTICATION_ARN

El ARN del certificado de AWS Certificate Manager utilizado por Amazon Redshift para la autenticación mtls con Apache Kafka, Confluent Cloud o Amazon Managed Streaming para Apache Kafka (Amazon MSK). El ARN está disponible en la consola de ACM al elegir el certificado emitido.

SECRET_ARN

El nombre de recurso de Amazon (ARN) o un secreto compatible creado con AWS Secrets Manager. Para obtener información sobre cómo crear y recuperar un ARN para un secreto, consulte Administrar secretos con AWS Secrets Manager en la Guía del usuario de AWS Secrets Manager, y Recuperar el nombre de recurso de Amazon (ARN) del secreto en Amazon Redshift.

URI

La URL de arranque del clúster de Apache Kafka, Confluent Cloud o Amazon Managed Streaming para Apache Kafka (Amazon MSK). El punto de conexión debe ser accesible (enrutable) desde el clúster de Amazon Redshift. Para obtener más información, consulte CREATE EXTERNAL SCHEMA.

## ALTER EXTERNAL VIEW

Utilice el comando ALTER EXTERNAL VIEW para actualizar la vista externa. Según los parámetros que utilice, es posible que también se vean afectados otros motores de SQL, como Amazon Athena y Amazon EMR Spark, que también pueden hacer referencia a esta vista. Para obtener más información sobre las vistas del Catálogo de datos, consulte Vistas de AWS Glue Data Catalog.

## Sintaxis

```
ALTER EXTERNAL VIEW schema_name.view_name
{catalog_name.schema_name.view_name | awsdatacatalog.dbname.view_name |
 external_schema_name.view_name}
[FORCE] { AS (query_definition) | REMOVE DEFINITION }
```

## Parámetros

schema_name.view_name

Es el esquema asociado a la base de datos de AWS Glue, seguido del nombre de la vista.

catalog_name.schema_name.view_name | awsdatacatalog.dbname.view_name | external_schema_name.view_name

Es la notación del esquema que se utilizará al modificar la vista. Puede especificar el uso de AWS Glue Data Catalog, una base de datos de Glue que haya creado o un esquema externo que haya creado. Consulte CREATE DATABASE y CREATE EXTERNAL SCHEMA para obtener más información.

FORCE

Si AWS Lake Formation debe actualizar la definición de la vista incluso si los objetos a los que se hace referencia en la tabla no son coherentes con otros motores de SQL. Si Lake Formation actualiza la vista, esta se considera obsoleta para los demás motores de SQL hasta que dichos motores se actualicen.

AS query_definition

Es la definición de la consulta SQL que Amazon Redshift ejecuta para modificar la vista.

REMOVE DEFINITION

Si se deben eliminar las vistas y volver a crearlas. Las vistas deben suprimirse y volver a crearse para marcarlas como `PROTECTED`.

## Ejemplos

El siguiente ejemplo modifica una vista del catálogo de datos denominada sample_schema.glue_data_catalog_view.

```
ALTER EXTERNAL VIEW sample_schema.glue_data_catalog_view
FORCE
REMOVE DEFINITION
```

# ALTER FUNCTION

Cambia el nombre de una función o cambia el propietario. Se requieren tanto el nombre de la función como los tipos de datos. Solo el propietario o un superusuario pueden cambiar el nombre de una función. Solo un superusuario puede cambiar el propietario de una función.

## Sintaxis

```
ALTER FUNCTION function_name ( { [ py_arg_name py_arg_data_type | sql_arg_data_type }
 [ , ... ] ] )
    RENAME TO new_name
```

```
ALTER FUNCTION function_name ( { [ py_arg_name py_arg_data_type | sql_arg_data_type }
 [ , ... ] ] )
    OWNER TO { new_owner | CURRENT_USER | SESSION_USER }
```

## Parámetros

function_name

El nombre de la función que se va a alterar. Especifique el nombre de la función en la ruta de búsqueda actual o utilice el formato `schema_name.function_name` para usar un esquema específico.

py_arg_name py_arg_data_type | sql_arg_data_type

Opcional. Una lista de nombres de argumentos de entrada y tipos de datos para la función definida por el usuario de Python o una lista de tipos de datos de argumentos de entrada para la función SQL definida por el usuario.

new_name

Un nuevo nombre para la función definida por el usuario.

new_owner | CURRENT_USER | SESSION_USER

Un nuevo propietario para la función definida por el usuario.

### Ejemplos

El siguiente ejemplo cambia el nombre de una función de `first_quarter_revenue` a `quarterly_revenue`.

```
ALTER FUNCTION first_quarter_revenue(bigint, numeric, int)
         RENAME TO quarterly_revenue;
```

En el siguiente ejemplo, se modifica el propietario de la función `quarterly_revenue` a `etl_user`.

```
ALTER FUNCTION quarterly_revenue(bigint, numeric) OWNER TO etl_user;
```

## ALTER GROUP

Cambia un grupo de usuarios. Utilice este comando para agregar usuarios al grupo, eliminar usuarios del grupo o cambiar el nombre del grupo.

### Sintaxis

```
ALTER GROUP group_name
{
ADD USER username [, ... ] |
DROP USER username [, ... ] |
RENAME TO new_name
}
```

### Parámetros

group_name

Nombre del grupo de usuarios que se modificará.

ADD

Agrega un usuario a un grupo de usuarios.

DROP

Elimina un usuario del grupo de usuarios.

username

Nombre del usuario que se agregará al grupo o que se eliminará del grupo.

RENAME TO

Cambia el nombre del grupo de usuarios. Los nombres de grupos que comienzan con dos guiones bajos se reservan para el uso interno de Amazon Redshift. Para obtener más información acerca de los nombres válidos, consulte Nombres e identificadores.

new_name

Nuevo nombre del grupo de usuarios.

### Ejemplos

En el siguiente ejemplo, se agrega un usuario denominado DWUSER al grupo ADMIN_GROUP.

```
ALTER GROUP admin_group
ADD USER dwuser;
```

En el siguiente ejemplo, se cambia el nombre del grupo ADMIN_GROUP a ADMINISTRATORS.

```
ALTER GROUP admin_group
RENAME TO administrators;
```

En el siguiente ejemplo, se agrega un usuario denominado DWUSER al grupo ADMIN_GROUP.

```
ALTER GROUP admin_group
ADD USER u1, u2;
```

En el siguiente ejemplo se eliminan dos usuarios del grupo ADMIN_GROUP.

```
ALTER GROUP admin_group
DROP USER u1, u2;
```

## MODIFICAR PROVEEDOR DE IDENTIDADES

Modifica un proveedor de identidades para asignar nuevos parámetros y valores. Cuando ejecuta este comando, todos los valores de parámetros establecidos anteriormente se eliminan antes de asignarse los nuevos valores. Solo un superusuario puede modificar un proveedor de identidades.

### Sintaxis

```
ALTER IDENTITY PROVIDER identity_provider_name
```

```
[PARAMETERS parameter_string]
[NAMESPACE namespace]
[IAM_ROLE iam_role]
[AUTO_CREATE_ROLES
    [ TRUE [ { INCLUDE | EXCLUDE } GROUPS LIKE filter_pattern] |
      FALSE
    ]
[DISABLE | ENABLE]
```

## Parámetros

identity_provider_name

Nombre del nuevo proveedor de identidades. Para obtener más información acerca de los nombres válidos, consulte Nombres e identificadores.

parameter_string

Una cadena que contiene un objeto JSON con el formato correcto que contiene los parámetros y valores necesarios para el proveedor de identidades específico.

namespace

El espacio de nombres de la organización.

iam_role

El rol de IAM que proporciona permisos para la conexión al IAM Identity Center. Este parámetro solo se aplica cuando el tipo de proveedor de identidades es AWSIDC.

auto_create_roles

Habilita o deshabilita la característica de creación automática de roles. Si el valor es TRUE, Amazon Redshift habilita la característica de creación automática de roles. Si el valor es FALSE, Amazon Redshift desactiva la característica de creación automática de roles. Si no se especifica el valor de este parámetro, Amazon Redshift determina el valor mediante la siguiente lógica:

- Si se proporciona `AUTO_CREATE_ROLES` pero no se especifica el valor, el valor se establece en TRUE.
- Si no se proporciona `AUTO_CREATE_ROLES` y el proveedor de identidad es AWSIDC, el valor se establece en FALSE.
- Si no se proporciona `AUTO_CREATE_ROLES` y el proveedor de identidad es Azure, el valor se establece en TRUE.

Para incluir grupos, especifique `INCLUDE`. El valor predeterminado es vacío, lo que significa incluir todos los grupos si `AUTO_CREATE_ROLES` está activado.

Para excluir grupos, especifique `EXCLUDE`. El valor predeterminado es vacío, lo que significa que no se excluirá ningún grupo si `AUTO_CREATE_ROLES` está activado.

filter_pattern

Una expresión de caracteres UTF-8 válida con un patrón para hacer coincidir los nombres de grupo. La opción LIKE realiza una coincidencia que distingue entre mayúsculas y minúsculas y admite los siguientes metacaracteres de coincidencia de patrones:

| Metacarácter | Descripción |
|---|---|
| % | Coincide con cualquier secuencia de cero o más caracteres. |
| _ | Coincide con cualquier carácter. |

Si filter_pattern no contiene metacaracteres, solo representa la propia cadena; en ese caso, LIKE actúa igual que el operador de igualdad.

filter_pattern admite los siguientes caracteres:

- Caracteres alfabéticos en mayúsculas y minúsculas (A-Z y a-z)
- Números (0-9)
- Los siguientes caracteres especiales:

```
_ % ^ * + ? { } , $
```

DISABLE o ENABLE

Activa o desactiva un proveedor de identidades. El valor predeterminado es ENABLE.

## Ejemplos

En el siguiente ejemplo se modifica un proveedor de identidades denominado oauth_standard. Se aplica específicamente a los casos en que Microsoft Azure AD es el proveedor de identidades.

```
ALTER IDENTITY PROVIDER oauth_standard
```

```
PARAMETERS '{"issuer":"https://sts.windows.net/2sdfdsf-d475-420d-b5ac-667adad7c702/",
"client_id":"87f4aa26-78b7-410e-bf29-57b39929ef9a",
"client_secret":"BUAH~ewrqewrqwerUUY^%tHe1oNZShoiU7",
"audience":["https://analysis.windows.net/powerbi/connector/AmazonRedshift"]
}'
```

El siguiente ejemplo muestra cómo configurar el espacio de nombres del proveedor de identidades. Esto se puede aplicar a Microsoft Azure AD, si sigue una instrucción como la del ejemplo anterior o a otro proveedor de identidades. También se puede aplicar a los casos en los que conecte un clúster aprovisionado de Amazon Redshift o un grupo de trabajo de Amazon Redshift sin servidor existente a IAM Identity Center, si tiene una conexión configurada a través de una aplicación administrada.

```
ALTER IDENTITY PROVIDER "my-redshift-idc-application"
NAMESPACE 'MYCO';
```

El siguiente ejemplo establece el rol de IAM y funciona en el caso de uso para configurar la integración de Redshift con IAM Identity Center.

```
ALTER IDENTITY PROVIDER "my-redshift-idc-application"
IAM_ROLE 'arn:aws:iam::123456789012:role/myadministratorrole';
```

Para obtener más información sobre la configuración de una conexión de IAM Identity Center desde Redshift, consulte Conectar Redshift con IAM Identity Center para ofrecer a los usuarios una experiencia de inicio de sesión único.

Desactivación de un proveedor de identidades

La instrucción de ejemplo siguiente muestra cómo desactivar un proveedor de identidades. Cuando está desactivado, los usuarios federados del proveedor de identidades no pueden iniciar sesión en el clúster hasta que se vuelva a habilitar.

```
ALTER IDENTITY PROVIDER "redshift-idc-app" DISABLE;
```

## ALTER MASKING POLICY

Modifica una política de enmascaramiento de datos dinámicos existente. Para obtener más información sobre el enmascaramiento dinámico de datos, consulte Enmascaramiento de datos dinámico.

Los superusuarios y los usuarios o roles que tienen el rol sys:secadmin pueden modificar una política de enmascaramiento.

### Sintaxis

```
ALTER MASKING POLICY policy_name
   USING (masking_expression);
```

### Parámetros

policy_name

Nombre de la política de enmascaramiento. Debe ser el nombre de una política de enmascaramiento que ya exista en la base de datos.

masking_expression

Expresión SQL que se utiliza para transformar las columnas de destino. Se puede escribir mediante funciones de manipulación de datos, como las funciones de manipulación de cadenas, o junto con funciones definidas por el usuario escritas en SQL, Python o con AWS Lambda.

La expresión debe coincidir con las columnas de entrada y los tipos de datos de la expresión original. Por ejemplo, si las columnas de entrada de la política de enmascaramiento original fueran `sample_1 FLOAT` y`sample_2 VARCHAR(10)`, no podría modificar la política de enmascaramiento para que ocupara una tercera columna ni hacer que la política aceptara un valor FLOAT y un valor BOOLEAN. Si usa una constante como expresión de enmascaramiento, debe convertirla explícitamente a un tipo que coincida con el tipo de entrada.

Debe tener el permiso USAGE en todas las funciones definidas por el usuario que utilice en la expresión de enmascaramiento.

## ALTER MATERIALIZED VIEW

Cambia los atributos de una vista materializada.

### Sintaxis

```
ALTER MATERIALIZED VIEW mv_name
{
AUTO REFRESH { YES | NO }
```

```
| ALTER DISTKEY column_name
| ALTER DISTSTYLE ALL
| ALTER DISTSTYLE EVEN
| ALTER DISTSTYLE KEY DISTKEY column_name
| ALTER DISTSTYLE AUTO
| ALTER [COMPOUND] SORTKEY ( column_name [,...] )
| ALTER SORTKEY AUTO
| ALTER SORTKEY NONE
| ROW LEVEL SECURITY { ON | OFF } [ CONJUNCTION TYPE { AND | OR } ] [FOR DATASHARES]
};
```

## Parámetros

mv_name

Se trata del nombre de la vista materializada que se modificará.

AUTO REFRESH { YES | NO }

Una cláusula que activa o desactiva la actualización automática de una vista materializada. Para obtener más información acerca de la actualización automática de vistas materializadas, consulte Actualización de una vista materializada.

ALTER DISTSTYLE ALL

Cláusula que cambia el estilo de distribución existente de una relación con ALL. Considere lo siguiente:

- Las operaciones ALTER DISTSTYTLE, ALTER SORTKEY y VACUUM no pueden ejecutarse simultáneamente en la misma relación.
  - Si se está ejecutando la operación VACUUM actualmente, la ejecución de ALTER DISTSTYLE ALL devuelve un error.
  - Si se está ejecutando ALTER DISTSTYLE ALL, no se iniciará una limpieza en segundo plano en una relación.
- El comando ALTER DISTSTYLE ALL no está admitido en relaciones con claves de clasificación intercaladas y tablas temporales.
- Si el estilo de distribución se definió previamente como AUTO, la relación ya no será candidata para la optimización automática de tablas.

Para obtener más información acerca de DISTSTYLE ALL, vaya a CREATE MATERIALIZED VIEW.

ALTER DISTSTYLE EVEN

Cláusula que cambia el estilo de distribución existente de una relación con EVEN. Considere lo siguiente:

- Las operaciones ALTER DISTSYTLE, ALTER SORTKEY y VACUUM no pueden ejecutarse simultáneamente en la misma relación.
  - Si se está ejecutando VACUUM actualmente, la ejecución de ALTER DISTSTYLE EVEN devuelve un error.
  - Si se está ejecutando ALTER DISTSTYLE EVEN, no se iniciará una limpieza en segundo plano en una relación.
- El comando ALTER DISTSTYLE EVEN no está admitido en relaciones con claves de clasificación intercaladas y tablas temporales.
- Si el estilo de distribución se definió previamente como AUTO, la relación ya no será candidata para la optimización automática de tablas.

Para obtener más información acerca de DISTSTYLE EVEN, consulte CREATE MATERIALIZED VIEW.

ALTER DISTKEY column_name o ALTER DISTSTYLE KEY DISTKEY column_name

Cláusula que cambia la columna usada como la clave de distribución de una relación. Considere lo siguiente:

- Las operaciones VACUUM y ALTER DISTKEY no se pueden ejecutar simultáneamente en la misma relación.
  - Si ya se está ejecutando VACUUM, ALTER DISTKEY devolverá un error.
  - Si se está ejecutando ALTER DISTKEY, la limpieza en segundo plano no se iniciará en una relación.
  - Si se está ejecutando ALTER DISTKEY, la limpieza en primer plano devolverá un error.
- Solo puede ejecutar un comando ALTER DISTKEY en una relación a la vez.
- El comando ALTER DISTKEY no es compatible en relaciones con claves de clasificación intercalada.
- Si el estilo de distribución se definió previamente como AUTO, la relación ya no será candidata para la optimización automática de tablas.

Al especificar DISTSTYLE KEY, los datos se distribuyen por los valores en la columna DISTKEY. Para obtener más información acerca de DISTSTYLE, consulte CREATE MATERIALIZED VIEW.

ALTER DISTSTYLE AUTO

Cláusula que cambia el estilo de distribución existente de una relación a AUTO.

Si modifica un estilo de distribución a AUTO, el estilo de distribución de la tabla se establece de la siguiente manera:

- Una relación pequeña con DISTSTYLE ALL se convierte a AUTO(ALL).
- Una relación pequeña con DISTSTYLE EVEN se convierte a AUTO(ALL).
- Una relación pequeña con DISTSTYLE KEY se convierte a AUTO(ALL).
- Una relación grande con DISTSTYLE ALL se convierte a AUTO(EVEN).
- Una relación grande con DISTSTYLE EVEN se convierte a AUTO(EVEN).
- Una relación grande con DISTSTYLE KEY se convierte a AUTO(KEY) y se conserva DISTKEY. En este caso, Amazon Redshift no cambia la relación.

Si Amazon Redshift determina que un estilo de distribución nuevo o una clave nueva mejorarán el rendimiento de las consultas, Amazon Redshift podría cambiar el estilo de distribución o la clave de la relación en el futuro. Por ejemplo, Amazon Redshift podría convertir una relación con un valor de DISTSTYLE de AUTO(KEY) en AUTO(EVEN) o viceversa. Para obtener más información sobre el comportamiento cuando se modifican las claves de distribución, incluida la redistribución de datos y los bloqueos, consulte Recomendaciones de Amazon Redshift Advisor.

Para obtener más información acerca de DISTSTYLE AUTO, consulte CREATE MATERIALIZED VIEW.

Para ver el estilo de distribución de una relación, consulte la vista de catálogo del sistema SVV_TABLE_INFO. Para obtener más información, consulte SVV_TABLE_INFO. Si desea ver las recomendaciones para relaciones de Advisor de Amazon Redshift, consulte la vista de catálogo del sistema SVV_ALTER_TABLE_RECOMMENDATIONS. Para obtener más información, consulte SVV_ALTER_TABLE_RECOMMENDATIONS. Para ver las acciones llevadas a cabo por Amazon Redshift, consulte la vista de catálogo del sistema SVL_AUTO_WORKER_ACTION. Para obtener más información, consulte SVL_AUTO_WORKER_ACTION.

ALTER [COMPOUND] SORTKEY ( column_name [,...] )

Una cláusula que cambia o añade la clave de clasificación utilizada para una relación. ALTER SORTKEY no se admite en tablas temporales.

Cuando se modifica una clave de ordenación, la codificación de compresión de las columnas de la clave de ordenación nueva u original puede cambiar. Si no se define de forma explícita

ninguna codificación para la relación, entonces Amazon Redshift asigna automáticamente las codificaciones de compresión de la siguiente manera:

- A las columnas que están definidas como claves de ordenación se les asigna una compresión RAW.
- A las columnas que están definidas como tipos de datos BOOLEAN, REAL o DOUBLE PRECISION se les asigna una compresión RAW.
- A las columnas que se definen como SMALLINT, INTEGER, BIGINT, DECIMAL, DATE, TIME, TIMETZ, TIMESTAMP o TIMESTAMPTZ se les asigna la compresión AZ64.
- Las columnas que se definen como CHAR o VARCHAR tienen asignada la compresión LZO.

Considere lo siguiente:

- Puede definir un máximo de 400 columnas para una clave de clasificación para cada relación.
- Puede alterar una clave de ordenación intercalada por una clave de ordenación compuesta o por ninguna clave de ordenación. No obstante, no puede alterar una clave de ordenación compuesta por una clave de ordenación intercalada.
- Si la clave de clasificación se definió previamente como AUTO, la relación ya no será candidata para la optimización automática de tablas.
- Amazon Redshift recomienda utilizar la codificación RAW (sin compresión) para columnas definidas como claves de ordenación. Cuando se modifica una columna para elegirla como clave de ordenación, la compresión de la columna se cambia a la compresión RAW (sin compresión). Esto puede aumentar la cantidad de almacenamiento que requiere la relación. El aumento del tamaño de la relación depende de la definición específica de la relación y de su contenido. Para obtener más información acerca de la compresión, consulte Codificaciones de compresión.

Cuando se cargan datos en una relación, se hace en el orden definido por la clave de clasificación. Cuando se modifica la clave de ordenación, Amazon Redshift reordena los datos. Para obtener más información acerca de SORTKEY, consulte CREATE MATERIALIZED VIEW.

ALTER SORTKEY AUTO

Se trata de una cláusula que modifica la clave de clasificación de la relación de destino a AUTO o la añade. ALTER SORTKEY AUTO no se admite en tablas temporales.

Cuando se modifica una clave de clasificación a AUTO, Amazon Redshift conserva la clave de ordenación existente de la relación.

Si Amazon Redshift determina que una clave de clasificación nueva mejorará el rendimiento de las consultas, Amazon Redshift podría cambiar la clave de clasificación de la relación en el futuro.

Para obtener más información acerca de SORTKEY AUTO, consulte CREATE MATERIALIZED VIEW.

Para ver la clave de clasificación de una relación, consulte la vista de catálogo del sistema SVV_TABLE_INFO. Para obtener más información, consulte SVV_TABLE_INFO. Si desea ver las recomendaciones para relaciones de Advisor de Amazon Redshift, consulte la vista de catálogo del sistema SVV_ALTER_TABLE_RECOMMENDATIONS. Para obtener más información, consulte SVV_ALTER_TABLE_RECOMMENDATIONS. Para ver las acciones llevadas a cabo por Amazon Redshift, consulte la vista de catálogo del sistema SVL_AUTO_WORKER_ACTION. Para obtener más información, consulte SVL_AUTO_WORKER_ACTION.

ALTER SORTKEY NONE

Se trata de una cláusula que quita la clave de clasificación de la relación de destino.

Si la clave de clasificación se definió previamente como AUTO, la relación ya no será candidata para la optimización automática de tablas.

ROW LEVEL SECURITY { ON | OFF } [ CONJUNCTION TYPE { AND | OR } ] [ FOR DATASHARES ]

Es una cláusula que activa o desactiva la seguridad de nivel de fila para una relación.

Cuando la seguridad de nivel de fila está activada para una relación, solo puede leer las filas a las que la política de seguridad de nivel de fila le permite acceder. Si no hay ninguna política que le conceda acceso a la relación, no podrá ver ninguna fila de la relación. Solo los superusuarios y los usuarios o roles que tienen el rol `sys:secadmin` pueden establecer la cláusula ROW LEVEL SECURITY. Para obtener más información, consulte Seguridad de nivel básico.

- [ CONJUNCTION TYPE { AND | OR } ]

  Es una cláusula que le permite elegir el tipo de conjunción de la política de seguridad a nivel de fila para una relación. Si se asocian varias políticas de seguridad en el nivel de fila a una relación, puede combinarlas con la cláusula AND u OR. De forma predeterminada, Amazon Redshift combina las políticas de RLS con la cláusula AND. Los superusuarios, usuarios o roles que tienen el rol `sys:secadmin` pueden usar esta cláusula para definir el tipo de conjunción de la política de seguridad de nivel de fila para una relación. Para obtener más información, consulte Combinación de varias políticas por usuario.
- PARA RECURSOS COMPARTIDOS DE DATOS

Es una cláusula que determina si se puede acceder a una relación protegida por RLS a través de recursos compartidos de datos. De forma predeterminada, no se puede acceder a una relación protegida por RLS a través de un recurso compartido de datos. Un comando ALTER MATERIALIZED VIEW ROW LEVEL SECURITY ejecutado con esta cláusula solo afecta a la propiedad de accesibilidad de recurso compartido de datos de la relación. La propiedad ROW LEVEL SECURITY no ha cambiado.

Si hace que una relación protegida por RLS sea accesible a través de recursos compartidos de datos, la relación no tendrá seguridad de nivel de fila en la base de datos de recurso compartido de datos del lado del consumidor. La relación conserva su propiedad RLS del lado del productor.

## Ejemplos

En el siguiente ejemplo, se habilita la actualización automática de la vista materializada `tickets_mv`.

```
ALTER MATERIALIZED VIEW tickets_mv AUTO REFRESH YES
```

## Ejemplos de DISTSTYLE y SORTKEY para ALTER MATERIALIZED VIEW

En los ejemplos de este tema se muestra cómo realizar cambios en DISTSTYLE y SORTKEY mediante ALTER MATERIALIZED VIEW.

Las consultas de ejemplo siguientes muestran cómo modificar una columna DISTSTYLE KEY DISTKEY mediante una tabla base de ejemplo:

```
CREATE TABLE base_inventory(
  inv_date_sk int4 NOT NULL,
  inv_item_sk int4 NOT NULL,
  inv_warehouse_sk int4 NOT NULL,
  inv_quantity_on_hand int4
);

INSERT INTO base_inventory VALUES(1,1,1,1);

CREATE materialized VIEW inventory diststyle even AS SELECT * FROM base_inventory;
SELECT "table", diststyle FROM svv_table_info WHERE "table" = 'inventory';
```

```
ALTER materialized VIEW inventory ALTER diststyle KEY distkey inv_warehouse_sk;
SELECT "table", diststyle FROM svv_table_info WHERE "table" = 'inventory';

ALTER materialized VIEW inventory ALTER distkey inv_item_sk;
SELECT "table", diststyle FROM svv_table_info WHERE "table" = 'inventory';

DROP TABLE base_inventory CASCADE;
```

Modifique una vista materializada a DISTSTYLE ALL:

```
CREATE TABLE base_inventory(
  inv_date_sk int4 NOT NULL,
  inv_item_sk int4 NOT NULL,
  inv_warehouse_sk int4 NOT NULL,
  inv_quantity_on_hand int4
);

INSERT INTO base_inventory VALUES(1,1,1,1);

CREATE materialized VIEW inventory diststyle even AS SELECT * FROM base_inventory;
SELECT "table", diststyle FROM svv_table_info WHERE "table" = 'inventory';

ALTER MATERIALIZED VIEW inventory ALTER diststyle ALL;
SELECT "table", diststyle FROM svv_table_info WHERE "table" = 'inventory';

DROP TABLE base_inventory CASCADE;
```

Los siguientes comandos muestran ejemplos de ALTER MATERIALIZED VIEW SORTKEY, con una tabla base de ejemplo:

```
CREATE TABLE base_inventory (c0 int, c1 int);

INSERT INTO base_inventory VALUES(1,1);

CREATE materialized VIEW inventory interleaved sortkey(c0, c1) AS SELECT * FROM
 base_inventory;
SELECT "table", sortkey1 FROM svv_table_info WHERE "table" = 'inventory';

ALTER materialized VIEW inventory ALTER sortkey(c0, c1);
SELECT "table", diststyle, sortkey_num FROM svv_table_info WHERE "table" = 'inventory';

ALTER materialized VIEW inventory ALTER sortkey NONE;
SELECT "table", diststyle, sortkey_num FROM svv_table_info WHERE "table" = 'inventory';
```

```
ALTER materialized VIEW inventory ALTER sortkey(c0);
SELECT "table", diststyle, sortkey_num FROM svv_table_info WHERE "table" = 'inventory';

DROP TABLE base_inventory CASCADE;
```

# ALTER RLS POLICY

Modifique una política de seguridad de nivel de fila existente en una tabla.

Los superusuarios y los usuarios o roles que tienen el rol `sys:secadmin` pueden modificar una política.

## Sintaxis

```
ALTER RLS POLICY policy_name
USING ( using_predicate_exp );
```

## Parámetros

policy_name

El nombre de la política de .

USING ( using_predicate_exp )

Especifica un filtro que se aplica a la cláusula WHERE de la consulta. Amazon Redshift aplica un predicado de política antes de los predicados de usuario de la consulta. Por ejemplo, **`current_user = ‘joe’ and price > 10`** limita a Joe a ver solo registros con un precio superior a 10 USD.

La expresión tiene acceso a las variables declaradas en la cláusula WITH de la instrucción CREATE RLS POLICY que se utilizó para crear la política con el nombre policy_name.

## Ejemplos

En el siguiente ejemplo se modifica una política de RLS.

```
-- First create an RLS policy that limits access to rows where catgroup is 'concerts'.
CREATE RLS POLICY policy_concerts
WITH (catgroup VARCHAR(10))
USING (catgroup = 'concerts');
```

```
-- Then, alter the RLS policy to only show rows where catgroup is 'piano concerts'.
ALTER RLS POLICY policy_concerts
USING (catgroup = 'piano concerts');
```

# ALTER ROLE

Cambia el nombre de un rol o cambia el propietario. Para obtener una lista de roles definidos por el sistema de Amazon Redshift, consulte the section called “Roles definidos por el sistema de Amazon Redshift”.

## Permisos necesarios

Los siguientes permisos son necesarios para ALTER ROLE:

- Superusuario
- Usuarios con los permisos ALTER ROLE

## Sintaxis

```
ALTER ROLE role [ WITH ]
  { { RENAME TO role } | { OWNER TO user_name } }[, ...]
  [ EXTERNALID TO external_id ]
```

## Parámetros

role

Nombre del rol que se debe modificar.

RENAME TO

Nuevo nombre del rol.

OWNER TO user_name

Nuevo propietario del rol.

EXTERNALID TO external_id

Un nuevo ID externo para el rol, que está asociado con un proveedor de identidades. Para obtener más información, consulte Federación de proveedores de identidades nativos (IdP) para Amazon Redshift.

### Ejemplos

El siguiente ejemplo cambia el nombre de un rol de `sample_role1` a `sample_role2`.

```
ALTER ROLE sample_role1 RENAME TO sample_role2;
```

El siguiente ejemplo cambia el propietario del rol.

```
ALTER ROLE sample_role1 WITH OWNER TO user1
```

La sintaxis de ALTER ROLE es similar a la de ALTER PROCEDURE, que aparece a continuación.

```
ALTER PROCEDURE first_quarter_revenue(bigint, numeric) RENAME TO quarterly_revenue;
```

En el siguiente ejemplo, se modifica el propietario de un procedimiento a `etl_user`.

```
ALTER PROCEDURE quarterly_revenue(bigint, numeric) OWNER TO etl_user;
```

En el ejemplo siguiente se actualiza un rol `sample_role1` con un nuevo ID externo asociado a un proveedor de identidades.

```
ALTER ROLE sample_role1 EXTERNALID TO "XYZ456";
```

## ALTER PROCEDURE

Cambia el nombre de un procedimiento o cambia el propietario. Se necesitan ambos, el nombre del procedimiento y los tipos de datos, o la firma. Solo el propietario o un superusuario pueden cambiar el nombre de un procedimiento. Solo un superusuario puede cambiar el propietario de un procedimiento.

### Sintaxis

```
ALTER PROCEDURE sp_name [ ( [ [ argname ] [ argmode ] argtype [, ...] ] ) ]
    RENAME TO new_name
```

```
ALTER PROCEDURE sp_name [ ( [ [ argname ] [ argmode ] argtype [, ...] ] ) ]
```

```
    OWNER TO { new_owner | CURRENT_USER | SESSION_USER }
```

## Parámetros

sp_name

El nombre del procedimiento que debe modificarse. Especifique solo el nombre del procedimiento en la ruta de búsqueda actual o utilice el formato `schema_name.sp_procedure_name` para usar un esquema específico.

[argname] [argmode] argtype

Una lista de nombres de argumento, modos de argumento y tipos de datos. Solo se requieren los tipos de datos de entrada, que se utilizan para identificar el procedimiento almacenado. Además, puede proporcionar la firma completa utilizada para crear el procedimiento incluidos los parámetros de entrada y salida con sus modos.

new_name

Un nombre nuevo para el procedimiento almacenado.

new_owner | CURRENT_USER | SESSION_USER

Un usuario nuevo para el procedimiento almacenado.

## Ejemplos

En el siguiente ejemplo, se cambia el nombre de un procedimiento de `first_quarter_revenue` a `quarterly_revenue`.

```
ALTER PROCEDURE first_quarter_revenue(volume INOUT bigint, at_price IN numeric,
 result OUT int) RENAME TO quarterly_revenue;
```

Este ejemplo equivale a lo siguiente.

```
ALTER PROCEDURE first_quarter_revenue(bigint, numeric) RENAME TO quarterly_revenue;
```

En el siguiente ejemplo, se modifica el propietario de un procedimiento a `etl_user`.

```
ALTER PROCEDURE quarterly_revenue(bigint, numeric) OWNER TO etl_user;
```

# ALTER SCHEMA

Cambia la definición de un esquema existente. Utilice este comando para cambiar el nombre de un esquema o el propietario de un esquema. Por ejemplo, cambie el nombre de un esquema existente para conservar una copia de seguridad de ese esquema cuando planifique crear una nueva versión de ese esquema. Para obtener más información acerca de los esquemas, consulte CREATE SCHEMA.

Para ver las cuotas del esquema configuradas, consulte SVV_SCHEMA_QUOTA_STATE.

Para ver los registros en los que se superaron las cuotas del esquema, consulte STL_SCHEMA_QUOTA_VIOLATIONS.

## Privilegios necesarios

Los siguientes privilegios son necesarios para ALTER SCHEMA:

- Superusuario
- Usuarios con el privilegio ALTER SCHEMA
- Propietario del esquema

Al cambiar el nombre de un esquema, tenga en cuenta que los objetos que utilizan el nombre anterior, como los procedimientos almacenados o las vistas materializadas, deben actualizarse para utilizar el nuevo nombre.

## Sintaxis

```
ALTER SCHEMA schema_name
{
RENAME TO new_name |
OWNER TO new_owner |
QUOTA { quota [MB | GB | TB] | UNLIMITED }
}
```

## Parámetros

schema_name

El nombre del esquema de la base de datos que se modificará.

RENAME TO

Una cláusula que cambia el nombre del esquema.

new_name

El nuevo nombre del esquema. Para obtener más información acerca de los nombres válidos, consulte Nombres e identificadores.

OWNER TO

Una cláusula que cambia el propietario del esquema.

new_owner

El nuevo propietario del esquema.

QUOTA

La cantidad máxima de espacio en disco que puede utilizar el esquema especificado. Este espacio es el tamaño colectivo de todas las tablas en el esquema especificado. Amazon Redshift convierte el valor seleccionado en megabytes. Gigabytes es la unidad de medida predeterminada cuando no se especifica un valor.

Para obtener más información acerca de la configuración de las cuotas del esquema, consulte CREATE SCHEMA.

## Ejemplos

En el siguiente ejemplo, se cambia el nombre del esquema SALES a US_SALES.

```
alter schema sales
rename to us_sales;
```

En el siguiente ejemplo, se otorga la propiedad del esquema US_SALES al usuario DWUSER.

```
alter schema us_sales
owner to dwuser;
```

En el siguiente ejemplo, se cambia la cuota a 300 MB y se quita la cuota.

```
alter schema us_sales QUOTA 300 GB;
```

```
alter schema us_sales QUOTA UNLIMITED;
```

# ALTER SYSTEM

Cambia una opción de configuración en el sistema para el clúster de Amazon Redshift o el grupo de trabajo de Amazon Redshift sin servidor.

## Privilegios necesarios

Uno de los siguientes tipos de usuario puede ejecutar el comando ALTER SYSTEM:

- Superusuario
- Usuario administrador

## Sintaxis

```
ALTER SYSTEM SET system-level-configuration = {true| t | on | false | f | off}
```

## Parámetros

system-level-configuration

Configuración en el nivel del sistema. Valores válidos: `data_catalog_auto_mount` y `metadata_security`.

{true| t | on | false | f | off}

Un valor para activar o desactivar la configuración en el nivel de sistema. Un valor `true`, `t` o on indica que se active la configuración. Un valor `false`, `f` o `off` indica que se desactive la configuración.

## Notas de uso

Para un clúster aprovisionado, los cambios en `data_catalog_auto_mount` entran en vigor en el siguiente reinicio del clúster. Para obtener más información, consulte Reinicio de un clúster en la Guía de administración de Amazon Redshift.

En el caso de un grupo de trabajo sin servidor, los cambios en `data_catalog_auto_mount` no se aplican de forma inmediata.

## Ejemplos

En el ejemplo siguiente se activa el montaje automático de AWS Glue Data Catalog.

```
ALTER SYSTEM SET data_catalog_auto_mount = true;
```

En el ejemplo siguiente se activa la seguridad de los metadatos.

```
ALTER SYSTEM SET metadata_security = true;
```

Establecimiento de un espacio de nombres de identidad predeterminado

Este ejemplo es específico para trabajar con un proveedor de identidades. Puede integrar Redshift con IAM Identity Center y un proveedor de identidades para centralizar la administración de identidades para Redshift y otros servicios de AWS.

El siguiente ejemplo muestra cómo configurar el espacio de nombres de identidades predeterminado para el sistema. Si lo hace posteriormente, resultará más sencillo ejecutar las instrucciones GRANT y CREATE, ya que no es necesario incluir el espacio de nombres como prefijo para cada identidad.

```
ALTER SYSTEM SET default_identity_namespace = 'MYCO';
```

Tras ejecutar el comando, puede ejecutar instrucciones como las siguientes:

```
GRANT SELECT ON TABLE mytable TO alice;

GRANT UPDATE ON TABLE mytable TO salesrole;

CREATE USER bob password 'md50c983d1a624280812631c5389e60d48c';
```

El efecto de establecer el espacio de nombres de identidades predeterminado es que cada identidad no lo requiere como prefijo. En este ejemplo, `alice` se sustituye por `MYCO:alice`. Esto ocurre con cualquier identidad incluida. Para obtener más información sobre cómo usar un proveedor de identidades con Redshift, consulte Conectar Redshift con IAM Identity Center para ofrecer a los usuarios una experiencia de inicio de sesión único.

Para obtener más información acerca de los ajustes relacionados con la configuración de Redshift con IAM Identity Center, consulte SET y MODIFICAR PROVEEDOR DE IDENTIDADES.

# ALTER TABLE

Este comando cambia la definición de una tabla de Amazon Redshift o de una tabla externa de Amazon Redshift Spectrum. Este comando actualiza los valores y las propiedades establecidos por CREATE TABLE o CREATE EXTERNAL TABLE. Puede utilizar ALTER TABLE en una vista para la seguridad a nivel de fila (RLS).

No puede ejecutar ALTER TABLE en una tabla externa dentro de un bloque de transacción (BEGIN … END). Para obtener más información acerca de las transacciones, consulte Aislamiento serializable.

ALTER TABLE bloquea la tabla para operaciones de lectura y escritura hasta que se complete la transacción que incluye la operación ALTER TABLE, a menos que en la documentación se indique específicamente que se pueden realizar consultas de datos u otras operaciones en la tabla mientras se la modifica.

## Privilegios necesarios

El usuario que modifica una tabla necesita el privilegio adecuado para que el comando se ejecute correctamente. Según el comando ALTER TABLE, se requiere uno de los siguientes privilegios.

- Superusuario
- Usuarios con el privilegio ALTER TABLE
- Propietario de la tabla con el privilegio USAGE en el esquema

## Sintaxis

```
ALTER TABLE table_name
{
ADD table_constraint
| DROP CONSTRAINT constraint_name [ RESTRICT | CASCADE ]
| OWNER TO new_owner
| RENAME TO new_name
| RENAME COLUMN column_name TO new_name
| ALTER COLUMN column_name TYPE updated_varchar_data_type_size
| ALTER COLUMN column_name ENCODE new_encode_type
| ALTER COLUMN column_name ENCODE encode_type,
| ALTER COLUMN column_name ENCODE encode_type, .....;
| ALTER DISTKEY column_name
| ALTER DISTSTYLE ALL
```

```
| ALTER DISTSTYLE EVEN
| ALTER DISTSTYLE KEY DISTKEY column_name
| ALTER DISTSTYLE AUTO
| ALTER [COMPOUND] SORTKEY ( column_name [,...] )
| ALTER SORTKEY AUTO
| ALTER SORTKEY NONE
| ALTER ENCODE AUTO
| ADD [ COLUMN ] column_name column_type
  [ DEFAULT default_expr ]
  [ ENCODE encoding ]
  [ NOT NULL | NULL ]
  [ COLLATE { CASE_SENSITIVE | CASE_INSENSITIVE } ] |
| DROP [ COLUMN ] column_name [ RESTRICT | CASCADE ]
| ROW LEVEL SECURITY { ON | OFF } [ CONJUNCTION TYPE { AND | OR } ] [ FOR DATASHARES ]
| MASKING { ON | OFF } FOR DATASHARES }

where table_constraint is:

[ CONSTRAINT constraint_name ]
{ UNIQUE ( column_name [, ... ] )
| PRIMARY KEY ( column_name [, ... ] )
| FOREIGN KEY (column_name [, ... ] )
   REFERENCES  reftable [ ( refcolumn ) ]}

The following options apply only to external tables:

SET LOCATION { 's3://bucket/folder/' | 's3://bucket/manifest_file' }
| SET FILE FORMAT format |
| SET TABLE PROPERTIES ('property_name'='property_value')
| PARTITION ( partition_column=partition_value [, ...] )
  SET LOCATION { 's3://bucket/folder' |'s3://bucket/manifest_file' }
| ADD [IF NOT EXISTS]
    PARTITION ( partition_column=partition_value [, ...] ) LOCATION
 { 's3://bucket/folder' |'s3://bucket/manifest_file' }
    [, ... ]
| DROP PARTITION ( partition_column=partition_value [, ...] )
```

Para reducir el tiempo de ejecución del comando ALTER TABLE, puede combinar algunas cláusulas del comando ALTER TABLE.

Amazon Redshift admite las siguientes combinaciones de las cláusulas ALTER TABLE:

```
ALTER TABLE tablename ALTER SORTKEY (column_list), ALTER DISTKEY column_Id;
```

```
ALTER TABLE tablename ALTER DISTKEY column_Id, ALTER SORTKEY (column_list);
ALTER TABLE tablename ALTER SORTKEY (column_list), ALTER DISTSTYLE ALL;
ALTER TABLE tablename ALTER DISTSTYLE ALL, ALTER SORTKEY (column_list);
```

## Parámetros

table_name

El nombre de la tabla que se modificará. Especifique solo el nombre de la tabla o use el formato nombre_de_esquema.nombre_de_tabla para usar un esquema específico. Las tablas externas deben estar clasificadas por el nombre de un esquema externo. También puede especificar un nombre de vista si está utilizando la instrucción ALTER TABLE para cambiar el nombre de una vista o cambiar su propietario. La longitud máxima del nombre de la tabla es de 127 bytes; los nombres más largos se truncan en 127 bytes. Puede usar caracteres multibyte UTF-8 de hasta un máximo de cuatro bytes. Para obtener más información acerca de los nombres válidos, consulte Nombres e identificadores.

ADD table_constraint

Una cláusula que agrega la restricción especificada a la tabla. Para obtener descripciones de los valores table_constraint válidos, consulte CREATE TABLE.

> **Note**
>
> No se puede agregar una restricción de clave principal a una columna que se puede anular. Si la columna se creó originalmente con la restricción NOT NULL, puede agregar la restricción de clave principal.

DROP CONSTRAINT constraint_name

Una cláusula que elimina la restricción mencionada de la tabla. Para eliminar una restricción, especifique el nombre de la restricción, no el tipo de restricción. Para ver los nombres de restricciones de la tabla, ejecute la siguiente consulta.

```
select constraint_name, constraint_type
from information_schema.table_constraints;
```

RESTRICT

Una cláusula que elimina solo la restricción especificada. RESTRICT es una opción para DROP CONSTRAINT. RESTRICT no puede utilizarse con CASCADE.

CASCADE

Una cláusula que elimina la restricción especificada y todos los elementos que dependen de esa restricción. CASCADE es una opción para DROP CONSTRAINT. CASCADE no puede utilizarse con RESTRICT.

OWNER TO new_owner

Cláusula que cambia el propietario de la tabla (o la vista) al valor new_owner.

RENAME TO new_name

Cláusula que cambia el nombre de una tabla (o una vista) al valor especificado en new_name. La longitud máxima del nombre de la tabla es de 127 bytes; los nombres más largos se truncan en 127 bytes.

No se puede cambiar el nombre de una tabla permanente a un nombre que comienza por "#". El nombre de una tabla que comienza con "#" indica una tabla temporal.

No se puede cambiar el nombre de una tabla externa.

ALTER COLUMN nombre_de_columna TYPE tamaño_de_tipo_de_datos_varchar_actualizado

Una cláusula que cambia el tamaño de una columna definida como un tipo de datos VARCHAR. Esta cláusula solo admite la modificación del tamaño de un tipo de datos VARCHAR. Tenga en cuenta las siguientes restricciones:

- No puede modificar una columna con codificaciones de compresión BYTEDICT, RUNLENGTH, TEXT255 o TEXT32K.
- No puede disminuir el tamaño por debajo del tamaño máximo de los datos existentes.
- No puede modificar columnas con valores predeterminados.
- No puede modificar columnas con UNIQUE, PRIMARY KEY o FOREIGN KEY.
- No puede modificar las columnas dentro de un bloque de transacción (BEGIN … END) Para obtener más información acerca de las transacciones, consulte Aislamiento serializable.

ALTER COLUMN column_name ENCODE new_encode_type

Se trata de una cláusula que cambia la codificación de compresión de una columna. Si especifica la codificación de compresión para una columna, la tabla ya no se establece en ENCODE AUTO.

Para obtener información acerca de la codificación de compresión, consulte Compresión de columnas para reducir el tamaño de los datos almacenados.

Cuando se cambia la codificación de compresión de una columna, la tabla sigue disponible para realizar consultas.

Tenga en cuenta las siguientes restricciones:

- No puede modificar una columna para que tenga la misma codificación que la definida actualmente.
- No puede modificar la codificación de una columna en una tabla con una clave de ordenación intercalada.

ALTER COLUMN column_name ENCODE encode_type, ALTER COLUMN column_name ENCODE encode_type, .....;

Se trata de una cláusula que modifica la codificación de compresión de varias columnas en un solo comando. Para obtener información acerca de la codificación de compresión, consulte Compresión de columnas para reducir el tamaño de los datos almacenados.

Cuando se cambia la codificación de compresión de una columna, la tabla sigue disponible para realizar consultas.

Tenga en cuenta las siguientes restricciones:

- No puede modificar una columna con el mismo tipo de codificación o uno diferente varias veces con un solo comando.
- No puede modificar una columna para que tenga la misma codificación que la definida actualmente.
- No puede modificar la codificación de una columna en una tabla con una clave de ordenación intercalada.

ALTER DISTSTYLE ALL

Cláusula que cambia el estilo de distribución existente de una tabla a ALL. Considere lo siguiente:

- Las operaciones ALTER DISTSTYTLE, ALTER SORTKEY y VACUUM no pueden ejecutarse simultáneamente en la misma tabla.
  - Si se está ejecutando la operación VACUUM actualmente, la ejecución de ALTER DISTSTYLE ALL devuelve un error.
  - Si se está ejecutando ALTER DISTSTYLE ALL, una limpieza en segundo plano no se iniciará en las tablas.

- El comando ALTER DISTSTYLE ALL no está admitido en tablas con claves de ordenación intercaladas y tablas temporales.
- Si el estilo de distribución se definió previamente como AUTO, la tabla ya no será candidata para la optimización automática de tablas.

Para obtener más información acerca de DISTSTYLE ALL, consulte CREATE TABLE.

ALTER DISTSTYLE EVEN

Cláusula que cambia el estilo de distribución existente de una tabla a EVEN. Considere lo siguiente:

- Las operaciones ALTER DISTSYTLE, ALTER SORTKEY y VACUUM no pueden ejecutarse simultáneamente en la misma tabla.
  - Si se está ejecutando VACUUM actualmente, la ejecución de ALTER DISTSTYLE EVEN devuelve un error.
  - Si se está ejecutando ALTER DISTSTYLE EVEN, una limpieza en segundo plano no se inicia en una tabla.
- El comando ALTER DISTSTYLE EVEN no se admite en tablas con claves de ordenación intercaladas y tablas temporales.
- Si el estilo de distribución se definió previamente como AUTO, la tabla ya no será candidata para la optimización automática de tablas.

Para obtener más información acerca de DISTSTYLE EVEN, consulte CREATE TABLE.

ALTER DISTKEY column_name o ALTER DISTSTYLE KEY DISTKEY column_name

Una cláusula que cambia la columna usada como la clave de distribución de una tabla. Considere lo siguiente:

- Las operaciones VACUUM y ALTER DISTKEY no se pueden ejecutar simultáneamente en la misma tabla.
  - Si ya se está ejecutando VACUUM, ALTER DISTKEY devolverá un error.
  - Si se está ejecutando ALTER DISTKEY, la limpieza en segundo plano no se iniciará en las tablas.
  - Si se está ejecutando ALTER DISTKEY, la limpieza en primer plano devolverá un error.
- Solo puede ejecutar un comando ALTER DISTKEY en una tabla cada vez.
- El comando ALTER DISTKEY no es compatible en tablas con claves de ordenación intercalada.

- Si el estilo de distribución se definió previamente como AUTO, la tabla ya no será candidata para la optimización automática de tablas.

Al especificar DISTSTYLE KEY, los datos se distribuyen por los valores en la columna DISTKEY. Para obtener más información acerca de DISTSTYLE, consulte CREATE TABLE.

ALTER DISTSTYLE AUTO

Se trata de una cláusula que cambia el estilo de distribución existente de una tabla a AUTO.

Si modifica un estilo de distribución a AUTO, el estilo de distribución de la tabla se establece en las siguientes opciones:

- Una tabla pequeña con DISTSTYLE ALL se convierte a AUTO(ALL).
- Una tabla pequeña con DISTSTYLE EVEN se convierte a AUTO(ALL).
- Una tabla pequeña con DISTSTYLE KEY se convierte a AUTO(ALL).
- Una tabla grande con DISTSTYLE ALL se convierte a AUTO(EVEN).
- Una tabla grande con DISTSTYLE EVEN se convierte a AUTO(EVEN).
- Una tabla grande con DISTSTYLE KEY se convierte a AUTO(KEY) y se conserva DISTKEY. En este caso, Amazon Redshift no cambia la tabla.

Si Amazon Redshift determina que un estilo de distribución nuevo o una clave nueva mejorarán el rendimiento de las consultas, Amazon Redshift podría cambiar el estilo de distribución o la clave de la tabla en el futuro. Por ejemplo, Amazon Redshift podría convertir una tabla con un valor de DISTSTYLE de AUTO(KEY) en AUTO(EVEN) o viceversa. Para obtener más información sobre el comportamiento cuando se modifican las claves de distribución, incluida la redistribución de datos y los bloqueos, consulte las Recomendaciones de Amazon Redshift Advisor.

Para obtener más información acerca de DISTSTYLE AUTO, consulte CREATE TABLE.

Para ver el estilo de distribución de una tabla, consulte la vista de catálogo del sistema SVV_TABLE_INFO. Para obtener más información, consulte SVV_TABLE_INFO. Si desea ver las recomendaciones para tablas de Advisor de Amazon Redshift, consulte la vista de catálogo del sistema SVV_ALTER_TABLE_RECOMMENDATIONS. Para obtener más información, consulte SVV_ALTER_TABLE_RECOMMENDATIONS. Para ver las acciones llevadas a cabo por Amazon Redshift, consulte la vista de catálogo del sistema SVL_AUTO_WORKER_ACTION. Para obtener más información, consulte SVL_AUTO_WORKER_ACTION.

ALTER [COMPOUND] SORTKEY ( column_name [,...] )

Una cláusula que cambia o agrega la clave de ordenación utilizada para una tabla. ALTER SORTKEY no se admite en tablas temporales.

Cuando se modifica una clave de ordenación, la codificación de compresión de las columnas de la clave de ordenación nueva u original puede cambiar. Si no se define de forma explícita ninguna codificación para la tabla, entonces Amazon Redshift asigna automáticamente las codificaciones de compresión de la siguiente manera:

- A las columnas que están definidas como claves de ordenación se les asigna una compresión RAW.
- A las columnas que están definidas como tipos de datos BOOLEAN, REAL o DOUBLE PRECISION se les asigna una compresión RAW.
- A las columnas que se definen como SMALLINT, INTEGER, BIGINT, DECIMAL, DATE, TIME, TIMETZ, TIMESTAMP o TIMESTAMPTZ se les asigna la compresión AZ64.
- Las columnas que se definen como CHAR o VARCHAR tienen asignada la compresión LZO.

Considere lo siguiente:

- Puede definir un máximo de 400 columnas para una clave de ordenación para cada tabla.
- Puede alterar una clave de ordenación intercalada por una clave de ordenación compuesta o por ninguna clave de ordenación. No obstante, no puede alterar una clave de ordenación compuesta por una clave de ordenación intercalada.
- Si la clave de ordenación se definió previamente como AUTO, la tabla ya no será candidata para la optimización automática de tablas.
- Amazon Redshift recomienda utilizar la codificación RAW (sin compresión) para columnas definidas como claves de ordenación. Cuando se modifica una columna para elegirla como clave de ordenación, la compresión de la columna se cambia a la compresión RAW (sin compresión). Esto puede aumentar la cantidad de almacenamiento que requiere la tabla. El aumento del tamaño de la tabla depende de la definición y el contenido específicos de la tabla. Para obtener más información acerca de la compresión, consulte Codificaciones de compresión.

Cuando se cargan datos en un tabla, se cargan en el orden definido por la clave de ordenación. Cuando se modifica la clave de ordenación, Amazon Redshift reordena los datos. Para obtener más información acerca de SORTKEY, consulte CREATE TABLE.

ALTER SORTKEY AUTO

Se trata de una cláusula que modifica la clave de ordenación de la tabla de destino a AUTO o la agrega. ALTER SORTKEY AUTO no se admite en tablas temporales.

Cuando se modifica una clave de ordenación a AUTO, Amazon Redshift conserva la clave de ordenación existente de la tabla.

Si Amazon Redshift determina que una clave de ordenación nueva mejorará el rendimiento de las consultas, Amazon Redshift podría cambiar la clave de ordenación de la tabla en el futuro.

Para obtener más información acerca de SORTKEY AUTO, consulte CREATE TABLE.

Para ver la clave de ordenación de una tabla, consulte la vista de catálogo del sistema SVV_TABLE_INFO. Para obtener más información, consulte SVV_TABLE_INFO. Si desea ver las recomendaciones para tablas de Advisor de Amazon Redshift, consulte la vista de catálogo del sistema SVV_ALTER_TABLE_RECOMMENDATIONS. Para obtener más información, consulte SVV_ALTER_TABLE_RECOMMENDATIONS. Para ver las acciones llevadas a cabo por Amazon Redshift, consulte la vista de catálogo del sistema SVL_AUTO_WORKER_ACTION. Para obtener más información, consulte SVL_AUTO_WORKER_ACTION.

ALTER SORTKEY NONE

Se trata de una cláusula que quita la clave de ordenación de la tabla de destino.

Si la clave de ordenación se definió previamente como AUTO, la tabla ya no será candidata para la optimización automática de tablas.

ALTER ENCODE AUTO

Se trata de una cláusula que cambia el tipo de codificación de las columnas de la tabla de destino a AUTO. Cuando se cambia la codificación a AUTO, Amazon Redshift conserva el tipo de codificación existente de las columnas de la tabla. Luego, si Amazon Redshift determina que un nuevo tipo de codificación puede mejorar el rendimiento de la consulta, Amazon Redshift puede cambiar el tipo de codificación de las columnas de la tabla.

Si modifica una o más columnas para especificar una codificación, Amazon Redshift ya no ajustará de forma automática la codificación para todas las columnas de la tabla. Las columnas retienen la configuración de la codificación actual.

Las siguientes acciones no afectan al ajuste ENCODE AUTO de la tabla:

- cambiar el nombre de la tabla
- modificar el ajuste DISTSTYLE o SORTKEY de la tabla
- agregar o soltar una columna con un ajuste ENCODE
- utilizar la opción COMPUPDATE del comando COPY Para obtener más información, consulte Operaciones de carga de datos.

Para ver la codificación de una tabla, consulte la vista de catálogo del sistema SVV_TABLE_INFO. Para obtener más información, consulte SVV_TABLE_INFO.

RENAME COLUMN column_name TO new_name

Cláusula que cambia el nombre de una columna al valor especificado en new_name. La longitud máxima del nombre de la columna es de 127 bytes; los nombres más largos se truncan en 127 bytes. Para obtener más información acerca de los nombres válidos, consulte Nombres e identificadores.

ADD [ COLUMN ] column_name

Una cláusula que agrega una columna con el nombre especificado a la tabla. Solo se puede agregar una columna en cada instrucción ALTER TABLE.

No se puede agregar una columna que sea la clave de distribución (DISTKEY) o la clave de ordenación (SORTKEY) de la tabla.

No se puede usar un comando ALTER TABLE ADD COLUMN para modificar los siguientes atributos de tabla y de columna:

- UNIQUE
- PRIMARY KEY
- REFERENCES (clave externa)
- IDENTITY o GENERATED BY DEFAULT AS IDENTITY

La longitud máxima del nombre de la columna es de 127 bytes; los nombres más largos se truncan en 127 bytes. La cantidad máxima de columnas que se pueden definir en una única tabla es 1 600.

Las restricciones siguientes se aplican al agregar una columna a una tabla externa:

- No se puede agregar una columna a una tabla externa con las restricciones de columna DEFAULT, ENCODE, NOT NULL o NULL.

- No se pueden agregar columnas a una tabla externa definida con el formato de archivo AVRO.
- Si están habilitadas las pseudocolumnas, el número máximo de columnas que se pueden definir en una única tabla externa es de 1 598. Si las pseudocolumnas no están habilitadas, el número máximo de columnas que se pueden definir en una única tabla es de 1 600.

Para obtener más información, consulte CREATE EXTERNAL TABLE.

column_type

El tipo de datos de la columna que se agrega. Para las columnas CHAR y VARCHAR, puede usar la palabra clave MAX en lugar de declarar una longitud máxima. MAX establece la longitud máxima en 4 096 bytes para CHAR o 65 535 bytes para VARCHAR. El tamaño máximo del objeto GEOMETRY es 1 048 447 bytes.

Para obtener información acerca de los tipos de datos compatibles con Amazon Redshift, consulte Tipos de datos.

DEFAULT default_expr

Una cláusula que asigna un valor de datos predeterminado para la columna. El tipo de datos de default_expr debe coincidir con el tipo de datos de la columna. El valor DEFAULT debe ser una expresión sin variables. No se permiten subconsultas, referencias cruzadas a otras columnas de la tabla actual ni funciones definidas por el usuario.

El valor default_expr se utiliza en cualquier operación INSERT en la que no se especifica un valor para la columna. Si no se especifica un valor predeterminado, el valor predeterminado para la columna es nulo.

Si una operación COPY se encuentra con un campo nulo en una columna que tiene un valor DEFAULT y una restricción NOT NULL, el comando COPY inserta el valor de default_expr.

DEFAULT no se admite para las tablas externas.

ENCODE encoding

La codificación de compresión de una columna. Si no se especifica la codificación de compresión para ninguna columna de una tabla o si se especifica la opción ENCODE AUTO para la tabla, Amazon Redshift administra automáticamente la codificación de compresión de todas las columnas de la tabla de manera predeterminada.

Si se especifica la codificación de compresión para cualquier columna de la tabla o si no se especifica la opción ENCODE AUTO para esta, Amazon Redshift asigna automáticamente la

codificación de compresión a las columnas para las cuales no se especifica dicha codificación, como se indica a continuación:

- Por defecto, se asigna una compresión RAW a todas las columnas de tablas temporales.
- A las columnas que están definidas como claves de ordenación se les asigna una compresión RAW.
- A las columnas que están definidas como tipos de datos BOOLEAN, REAL, DOUBLE PRECISION, GEOMETRY o GEOGRAPHY se les asigna una compresión RAW.
- A las columnas que se definen como SMALLINT, INTEGER, BIGINT, DECIMAL, DATE, TIME, TIMETZ, TIMESTAMP o TIMESTAMPTZ se les asigna la compresión AZ64.
- A las columnas que se definen como CHAR, VARCHAR o VARBYTE se les asigna la compresión LZO.

> **Note**
>
> Si no desea que una columna se comprima, especifique explícitamente la codificación RAW.

Se admiten los siguientes compression encodings (p. 68):

- AZ64
- BYTEDICT
- DELTA
- DELTA32K
- LZO
- MOSTLY8
- MOSTLY16
- MOSTLY32
- RAW (sin comprimir)
- RUNLENGTH
- TEXT255
- TEXT32K
- ZSTD

ENCODE no se admite para las tablas externas.

NOT NULL | NULL

NOT NULL indica que la columna no puede contener valores nulos. NULL, el valor predeterminado, especifica que la columna acepta valores nulos.

NOT NULL y NULL no se admiten para las tablas externas.

COLLATE { CASE_SENSITIVE | CASE_INSENSITIVE }

Se trata de una cláusula que especifica si la búsqueda o la comparación de cadenas en la columna es CASE_SENSITIVE o CASE_INSENSITIVE. El valor predeterminado es el mismo que la configuración actual de distinción entre mayúsculas y minúsculas de la base de datos.

Para encontrar información de la intercalación de bases de datos, utilice el siguiente comando:

```
SELECT db_collation();

db_collation
----------------
 case_sensitive
(1 row)
```

DROP [ COLUMN ] column_name

El nombre de la columna que se debe eliminar de la tabla.

No se puede eliminar la última columna de una tabla. Una tabla debe tener como mínimo una columna.

No se puede eliminar una columna que sea la clave de distribución (DISTKEY) o la clave de ordenación (SORTKEY) de la tabla. El comportamiento predeterminado de DROP COLUMN es RESTRICT si la columna tiene objetos dependientes, como vista, clave principal, clave externa o restricción UNIQUE.

Las restricciones siguientes se aplican al quitar una columna de una tabla externa:

- No se puede eliminar una columna de una tabla externa si la columna se utiliza como partición.
- No se puede eliminar una columna de una tabla externa definida con el formato de archivo AVRO.
- RESTRICT y CASCADE no se tienen en cuenta para las tablas externas.

- No puede eliminar las columnas de la tabla de políticas a la que se hace referencia dentro de la definición de política a menos que elimine o desactive la política. Esto también se aplica cuando se especifica la opción CASCADE. Puede eliminar otras columnas de la tabla de políticas.

Para obtener más información, consulte CREATE EXTERNAL TABLE.

RESTRICT

Cuando se usa con DROP COLUMN, RESTRICT significa que la columna que se va a eliminar no se eliminará, en los siguientes casos:

- Si una vista definida hace referencia a la columna que desea eliminar
- Si una clave externa hace referencia a la columna
- Si la columna forma parte de una clave multiparte

RESTRICT no puede utilizarse con CASCADE.

RESTRICT y CASCADE no se tienen en cuenta para las tablas externas.

CASCADE

Cuando se usa con DROP COLUMN, elimina la columna especificada y todos los elementos que dependen de esa columna. CASCADE no puede utilizarse con RESTRICT.

RESTRICT y CASCADE no se tienen en cuenta para las tablas externas.

Las siguientes opciones se aplican solo a tablas externas.

SET LOCATION { 's3://bucket/folder/' | 's3://bucket/manifest_file' }

Se trata de la ruta a la carpeta de Amazon S3 que contiene los archivos de datos o un archivo de manifiesto que incluye una lista de rutas de objetos de Amazon S3. Los buckets deben estar en la misma región de AWS que el clúster de Amazon Redshift. Para obtener una lista de las regiones de AWS admitidas, consulte Limitaciones de Amazon Redshift Spectrum. Para obtener más información acerca del uso de un archivo de manifiesto, consulte LOCATION en la sección Parámetros de la documentación sobre el comando CREATE EXTERNAL TABLE.

SET FILE FORMAT format

El formato de archivo para los archivos de datos externos.

Los formatos válidos son los siguientes:

- AVRO
- PARQUET
- RCFILE
- SEQUENCEFILE
- TEXTFILE

SET TABLE PROPERTIES ( 'property_name'='property_value')

Cláusula que establece la definición de las propiedades de una tabla externa.

> **Note**
>
> Las propiedades de tabla distinguen entre mayúsculas y minúsculas.

'numRows'='row_count'

Una propiedad que establece el valor de numRows para la definición de la tabla. Para actualizar de forma explícita las estadísticas de una tabla externa, establezca la propiedad numRows de manera que indique el tamaño de la tabla. Amazon Redshift no analiza las tablas externas para generar las estadísticas de las tablas que el optimizador de consultas emplea a la hora de crear un plan de consulta. Si no se establecen las estadísticas de la tabla que corresponden a una tabla externa, Amazon Redshift crea un plan de ejecución de consultas. El plan se basa en la suposición de que las tablas externas son las tablas más grandes y las tablas locales son las tablas más pequeñas

'skip.header.line.count'='line_count'

Una propiedad que establece el número de filas que se omiten al principio de cada archivo de código fuente.

PARTITION ( partition_column=partition_value [, ...] SET LOCATION { 's3://bucket/folder' | 's3://bucket/manifest_file' }

Una cláusula que configura una nueva ubicación para una o más columnas de partición.

ADD [ IF NOT EXISTS ] PARTITION ( partition_column=partition_value [, ...] ) LOCATION { 's3://bucket/folder' | 's3://bucket/manifest_file' } [, ... ]

Una cláusula que agrega una o más particiones. Puede especificar múltiples cláusulas PARTITION mediante una única instrucción ALTER TABLE … ADD.

> **Note**
>
> Si usa el catálogo de AWS Glue, puede agregar hasta 100 particiones mediante una única instrucción ALTER TABLE.

La cláusula IF NOT EXISTS indica que si la partición especificada ya existe, el comando no debería realizar ningún cambio. También indica que el comando debe devolver un mensaje en el que se indique que la tabla existe, en lugar de terminar con un error. Esta cláusula es útil cuando se realiza scripting, para que el script no produzca un error si ALTER TABLE intenta agregar una partición que ya existe.

DROP PARTITION (partition_column=partition_value [, ...] )

Una cláusula que elimina la partición especificada. Eliminar una partición modifica solo los metadatos de la tabla externa. Los datos de Amazon S3 no se ven afectados.

ROW LEVEL SECURITY { ON | OFF } [ CONJUNCTION TYPE { AND | OR } ] [ FOR DATASHARES ]

Es una cláusula que activa o desactiva la seguridad de nivel de fila para una relación.

Cuando la seguridad de nivel de fila está activada para una relación, solo puede leer las filas a las que la política de seguridad de nivel de fila le permite acceder. Si no hay ninguna política que le conceda acceso a la relación, no podrá ver ninguna fila de la relación. Solo los superusuarios y los usuarios o roles que tienen el rol `sys:secadmin` pueden establecer la cláusula ROW LEVEL SECURITY. Para obtener más información, consulte Seguridad de nivel básico.

- [ CONJUNCTION TYPE { AND | OR } ]

  Es una cláusula que le permite elegir el tipo de conjunción de la política de seguridad a nivel de fila para una relación. Si se asocian varias políticas de seguridad en el nivel de fila a una relación, puede combinarlas con la cláusula AND u OR. De forma predeterminada, Amazon Redshift combina las políticas de RLS con la cláusula AND. Los superusuarios, usuarios o roles que tienen el rol `sys:secadmin` pueden usar esta cláusula para definir el tipo de conjunción de la política de seguridad de nivel de fila para una relación. Para obtener más información, consulte Combinación de varias políticas por usuario.

- PARA RECURSOS COMPARTIDOS DE DATOS

  Es una cláusula que determina si se puede acceder a una relación protegida por RLS a través de recursos compartidos de datos. De forma predeterminada, no se puede acceder a una relación protegida por RLS a través de un recurso compartido de datos. Un comando ALTER

TABLE ROW LEVEL SECURITY ejecutado con esta cláusula solo afecta a la propiedad de accesibilidad de recurso compartido de datos de la relación. La propiedad ROW LEVEL SECURITY no ha cambiado.

Si hace que una relación protegida por RLS sea accesible a través de recursos compartidos de datos, la relación no tendrá seguridad de nivel de fila en la base de datos de recurso compartido de datos del lado del consumidor. La relación conserva su propiedad RLS del lado del productor.

MASKING { ON | OFF } FOR DATASHARES

Es una cláusula que determina si se puede acceder a una relación protegida por DDM a través de recursos compartidos de datos. De forma predeterminada, no se puede acceder a una relación protegida por DDM a través de un recurso compartido de datos. Si hace que una relación protegida por DDM sea accesible a través de recursos compartidos de datos, la relación no tendrá protección con enmascaramiento en la base de datos de recurso compartido de datos del lado del consumidor. La relación conserva su propiedad de enmascaramiento del lado del productor. Solo los superusuarios y los usuarios o roles que tengan el rol `sys:secadmin` pueden establecer la cláusula MASKING FOR DATASHARES. Para obtener más información, consulte Enmascaramiento de datos dinámico.

## Ejemplos

Para ver ejemplos que muestran cómo utilizar el comando ALTER TABLE, consulte lo siguiente.

- Ejemplos de ALTER TABLE
- Ejemplos de ALTER EXTERNAL TABLE
- Ejemplos de ALTER TABLE ADD y DROP COLUMN

## Ejemplos de ALTER TABLE

En los siguientes ejemplos se muestra el uso básico del comando ALTER TABLE.

Cambiar el nombre de una tabla o una vista

El siguiente comando cambia el nombre de la tabla USERS a USERS_BKUP:

```
alter table users
rename to users_bkup;
```

También puede utilizar este tipo de comando para cambiar el nombre de una vista.

Cambiar el propietario de una tabla o vista

El siguiente comando cambia el propietario de la tabla VENUE al usuario DWUSER:

```
alter table venue
owner to dwuser;
```

Los siguientes comandos crean una vista y, luego, cambian su propietario:

```
create view vdate as select * from date;
alter table vdate owner to vuser;
```

Cambiar el nombre de una columna

El siguiente comando cambia el nombre de la columna VENUESEATS de la tabla VENUE a VENUESIZE:

```
alter table venue
rename column venueseats to venuesize;
```

Eliminar la restricción de una tabla

Para eliminar la restricción de una tabla, como clave principal, clave externa o restricción única, primero debe buscar el nombre interno de la restricción. Y, a continuación, especificar el nombre de la restricción en el comando ALTER TABLE. En el siguiente ejemplo, se buscan las restricciones de la tabla CATEGORY y, luego, se elimina la clave principal con el nombre `category_pkey`.

```
select constraint_name, constraint_type
from information_schema.table_constraints
where constraint_schema ='public'
and table_name = 'category';

constraint_name | constraint_type
----------------+----------------
category_pkey   | PRIMARY KEY

alter table category
drop constraint category_pkey;
```

Modificar una columna VARCHAR

Para ahorrar espacio de almacenamiento, puede definir inicialmente una tabla con columnas VARCHAR con el tamaño mínimo necesario para sus requisitos de datos actuales. Si más adelante, para dar cabida a cadenas más largas, puede modificar la tabla para aumentar el tamaño de la columna.

En el siguiente ejemplo, se cambia el tamaño de la columna EVENTNAME a VARCHAR(300).

```
alter table event alter column eventname type varchar(300);
```

Modificar una columna VARBYTE

Para ahorrar espacio de almacenamiento, puede definir inicialmente una tabla con columnas VARBYTE con el tamaño mínimo necesario para sus requisitos de datos actuales. Si más adelante, para dar cabida a cadenas más largas, puede modificar la tabla para aumentar el tamaño de la columna.

En el siguiente ejemplo, se cambia el tamaño de la columna EVENTNAME a VARBYTE(300).

```
alter table event alter column eventname type varbyte(300);
```

Modificar la codificación de compresión de una columna

Puede modificar la codificación de compresión de una columna. A continuación, encontrará un conjunto de ejemplos que demuestran este enfoque. La definición de la tabla para estos ejemplos es la siguiente.

```
create table t1(c0 int encode lzo, c1 bigint encode zstd, c2 varchar(16) encode lzo, c3
 varchar(32) encode zstd);
```

La siguiente instrucción modifica la codificación de compresión de la columna c0 de codificación LZO a AZ64.

```
alter table t1 alter column c0 encode az64;
```

La siguiente instrucción modifica la codificación de compresión de la columna c1 de codificación Zstandard a AZ64.

```
alter table t1 alter column c1 encode az64;
```

La siguiente instrucción modifica la codificación de compresión de la columna c2 de codificación LZO a codificación por diccionario de bytes.

```
alter table t1 alter column c2 encode bytedict;
```

La siguiente instrucción modifica la codificación de compresión de la columna c3 de codificación Zstandard a Runlength.

```
alter table t1 alter column c3 encode runlength;
```

Modificar una columna DISTSTYLE KEY DISTKEY

Los siguientes ejemplos muestran cómo cambiar DISTSTYLE y DISTKEY en una tabla.

Crear una tabla con un estilo de distribución EVEN. LA vista SVV_TABLE_INFO muestra que DISTSYLE es EVEN.

```
create table inventory(
  inv_date_sk int4 not null ,
  inv_item_sk int4 not null ,
  inv_warehouse_sk int4 not null ,
  inv_quantity_on_hand int4
) diststyle even;

Insert into inventory values(1,1,1,1);

select "table", "diststyle" from svv_table_info;

   table   |   diststyle
-----------+----------------
 inventory |     EVEN
```

Altere la tabla DISTKEY a `inv_warehouse_sk`. La vista SVV_TABLE_INFO muestra la columna `inv_warehouse_sk` como la clave de distribución resultante.

```
alter table inventory alter diststyle key distkey inv_warehouse_sk;

select "table", "diststyle" from svv_table_info;

   table   |       diststyle
-----------+-----------------------
```

```
 inventory | KEY(inv_warehouse_sk)
```

Altere la tabla DISTKEY a `inv_item_sk`. La vista SVV_TABLE_INFO muestra la columna `inv_item_sk` como la clave de distribución resultante.

```
alter table inventory alter distkey inv_item_sk;

select "table", "diststyle" from svv_table_info;

   table   |       diststyle
-----------+----------------------
 inventory | KEY(inv_item_sk)
```

Modificar una tabla a DISTSTYLE ALL

En los ejemplos siguientes se muestra cómo cambiar una tabla a DISTSTYLE ALL.

Crear una tabla con un estilo de distribución EVEN. LA vista SVV_TABLE_INFO muestra que DISTSYLE es EVEN.

```
create table inventory(
  inv_date_sk int4 not null ,
  inv_item_sk int4 not null ,
  inv_warehouse_sk int4 not null ,
  inv_quantity_on_hand int4
) diststyle even;

Insert into inventory values(1,1,1,1);

select "table", "diststyle" from svv_table_info;

   table   |   diststyle
-----------+---------------
 inventory |     EVEN
```

Cambie la tabla DISTSTYLE a ALL. La vista SVV_TABLE_INFO muestra el DISTSYTLE modificado.

```
alter table inventory alter diststyle all;

select "table", "diststyle" from svv_table_info;

   table   |   diststyle
```

```
-----------+----------------
 inventory |     ALL
```

Modificar una tabla SORTKEY

Puede modificar una tabla para que tenga una clave de ordenación compuesta o ninguna clave de ordenación.

En la siguiente definición de la tabla, la tabla `t1` se define con una sortkey intercalada.

```
create table t1 (c0 int, c1 int) interleaved sortkey(c0, c1);
```

El siguiente comando modifica la tabla de una clave de ordenación intercalada a una clave de ordenación compuesta.

```
alter table t1 alter sortkey(c0, c1);
```

El siguiente comando modifica la tabla para eliminar la clave de ordenación intercalada.

```
alter table t1 alter sortkey none;
```

En la siguiente definición de la tabla, la tabla `t1` se define con una columna `c0` como sortkey.

```
create table t1 (c0 int, c1 int) sortkey(c0);
```

El siguiente comando modifica la tabla `t1` a una clave de ordenación compuesta.

```
alter table t1 alter sortkey(c0, c1);
```

Modificar una tabla para pasarla a ENCODE AUTO

En el siguiente ejemplo, se muestra cómo modificar una tabla para pasarla a ENCODE AUTO.

A continuación, se muestra la definición de la tabla para este ejemplo. La columna `c0` se define con el tipo de codificación AZ64 y la columna `c1` se define con el tipo de codificación LZO.

```
create table t1(c0 int encode AZ64, c1 varchar encode LZO);
```

Para esta tabla, la siguiente instrucción modifica la codificación a AUTO.

```
alter table t1 alter encode auto;
```

En el siguiente ejemplo, se muestra cómo modificar una tabla para quitar el ajuste ENCODE AUTO.

A continuación, se muestra la definición de la tabla para este ejemplo. Las columnas de la tabla se definen sin codificación. En este caso, la codificación adopta la opción ENCODE AUTO de manera predeterminada.

```
create table t2(c0 int, c1 varchar);
```

Para esta tabla, la siguiente instrucción modifica la codificación de la columna c0 a LZO. La codificación de la tabla ya no se establece en ENCODE AUTO.

```
alter table t2 alter column c0 encode lzo;;
```

Modificar el control de seguridad de nivel de fila

El siguiente comando desactiva la RLS de la tabla:

```
ALTER TABLE tickit_category_redshift ROW LEVEL SECURITY OFF;
```

El siguiente comando activa la RLS para la tabla:

```
ALTER TABLE tickit_category_redshift ROW LEVEL SECURITY ON;
```

El siguiente comando activa la RLS para la tabla y la hace accesible a través de recursos compartidos de datos:

```
ALTER TABLE tickit_category_redshift ROW LEVEL SECURITY ON;
ALTER TABLE tickit_category_redshift ROW LEVEL SECURITY FOR DATASHARES OFF;
```

El siguiente comando activa la RLS para la tabla y la hace inaccesible a través de recursos compartidos de datos:

```
ALTER TABLE tickit_category_redshift ROW LEVEL SECURITY ON;
ALTER TABLE tickit_category_redshift ROW LEVEL SECURITY FOR DATASHARES ON;
```

El siguiente comando activa la RLS y establece el tipo de conjunción RLS en OR para la tabla:

```
ALTER TABLE tickit_category_redshift ROW LEVEL SECURITY ON CONJUNCTION TYPE OR;
```

El siguiente comando activa la RLS y establece el tipo de conjunción RLS en AND para la tabla:

```
ALTER TABLE tickit_category_redshift ROW LEVEL SECURITY ON CONJUNCTION TYPE AND;
```

## Ejemplos de ALTER EXTERNAL TABLE

En los siguientes ejemplos, se utiliza un bucket de Amazon S3 situado en la región Este de EE.UU. (Norte de Virginia) (`us-east-1`) Región de AWS y las tablas de ejemplo creadas en Ejemplos para CREATE TABLE. Para obtener más información acerca de cómo utilizar las particiones con tablas externas, consulte Partición de tablas externas de Redshift Spectrum.

En los siguientes conjuntos de ejemplos se configura la propiedad numRows de la tabla para la tabla externa SPECTRUM.SALES en 170 000 filas.

```
alter table spectrum.sales
set table properties ('numRows'='170000');
```

En el siguiente ejemplo, se cambia la ubicación de la tabla externa SPECTRUM.SALES.

```
alter table spectrum.sales
set location 's3://redshift-downloads/tickit/spectrum/sales/';
```

En el siguiente ejemplo, se cambia el formato de la tabla externa SPECTRUM.SALES a Parquet.

```
alter table spectrum.sales
set file format parquet;
```

En el siguiente ejemplo, se agrega una partición para la tabla SPECTRUM.SALES_PART.

```
alter table spectrum.sales_part
add if not exists partition(saledate='2008-01-01')
location 's3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-01/';
```

En el siguiente ejemplo, se agregan tres particiones para la tabla SPECTRUM.SALES_PART.

```
alter table spectrum.sales_part add if not exists
partition(saledate='2008-01-01')
```

```
location 's3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-01/'
partition(saledate='2008-02-01')
location 's3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-02/'
partition(saledate='2008-03-01')
location 's3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-03/';
```

En el siguiente ejemplo, se modifica SPECTRUM.SALES_PART para eliminar la partición con `saledate='2008-01-01''`.

```
alter table spectrum.sales_part
drop partition(saledate='2008-01-01');
```

En el siguiente ejemplo, se establece una ruta nueva de Amazon S3 para la partición con `saledate='2008-01-01'`.

```
alter table spectrum.sales_part
partition(saledate='2008-01-01')
set location 's3://redshift-downloads/tickit/spectrum/sales_partition/
saledate=2008-01-01/';
```

En el siguiente ejemplo, se cambia el nombre de `sales_date` a `transaction_date`.

```
alter table spectrum.sales rename column sales_date to transaction_date;
```

En el siguiente ejemplo, se establece una asignación de columna con una asignación de posición en una tabla externa que utiliza el formato ORC (Optimized Row Columnar).

```
alter table spectrum.orc_example
set table properties('orc.schema.resolution'='position');
```

En el siguiente ejemplo, se establece una asignación de columna con una asignación de nombre en una tabla externa que utiliza el formato ORC.

```
alter table spectrum.orc_example
set table properties('orc.schema.resolution'='name');
```

## Ejemplos de ALTER TABLE ADD y DROP COLUMN

En los siguientes ejemplos se muestra cómo usar ALTER TABLE para agregar y eliminar la columna de una tabla básica y, también, cómo eliminar una columna con un objeto dependiente.

Utilizar ADD y, luego, DROP en una columna básica

En el siguiente ejemplo, se agrega una columna FEEDBACK_SCORE independiente a la tabla USERS. Esta columna contiene un valor entero, y el valor predeterminado de esta columna es NULL (no hay marcador de retroalimentación).

En primer lugar, consulte la tabla de catálogos PG_TABLE_DEF para ver el esquema de la tabla USERS:

```
column        | type                   | encoding | distkey | sortkey
--------------+------------------------+----------+---------+--------
userid        | integer                | delta    | true    |       1
username      | character(8)           | lzo      | false   |       0
firstname     | character varying(30)  | text32k  | false   |       0
lastname      | character varying(30)  | text32k  | false   |       0
city          | character varying(30)  | text32k  | false   |       0
state         | character(2)           | bytedict | false   |       0
email         | character varying(100) | lzo      | false   |       0
phone         | character(14)          | lzo      | false   |       0
likesports    | boolean                | none     | false   |       0
liketheatre   | boolean                | none     | false   |       0
likeconcerts  | boolean                | none     | false   |       0
likejazz      | boolean                | none     | false   |       0
likeclassical | boolean                | none     | false   |       0
likeopera     | boolean                | none     | false   |       0
likerock      | boolean                | none     | false   |       0
likevegas     | boolean                | none     | false   |       0
likebroadway  | boolean                | none     | false   |       0
likemusicals  | boolean                | none     | false   |       0
```

Ahora, agregue la columna feedback_score:

```
alter table users
add column feedback_score int
default NULL;
```

Seleccione la columna FEEDBACK_SCORE de USERS para verificar que se haya agregado:

```
select feedback_score from users limit 5;

feedback_score
---------------
```

```
NULL
NULL
NULL
NULL
NULL
```

Arrastre la columna para restablecer la DDL original:

```
alter table users drop column feedback_score;
```

Eliminación de una columna con un objeto dependiente

En el siguiente ejemplo, se elimina una columna que tiene un objeto dependiente. Como resultado, también se elimina el objeto dependiente.

Para comenzar, agregue la columna FEEDBACK_SCORE a la tabla USERS nuevamente:

```
alter table users
add column feedback_score int
default NULL;
```

Luego, cree una vista de la tabla USERS denominada USERS_VIEW:

```
create view users_view as select * from users;
```

Ahora, pruebe eliminar la columna FEEDBACK_SCORE de la tabla USERS. Esta instrucción DROP usa el comportamiento predeterminado (RESTRICT):

```
alter table users drop column feedback_score;
```

Amazon Redshift muestra un mensaje de error en el que se indica que la columna no puede eliminarse porque otro objeto depende de ella.

Pruebe eliminar la columna FEEDBACK_SCORE nuevamente y, esta vez, especifique CASCADE para eliminar todos los objetos dependientes:

```
alter table users
drop column feedback_score cascade;
```

# ALTER TABLE APPEND

Adjunta filas en una tabla de destino al mover datos de una tabla de origen existente. Los datos de la tabla de origen se mueven a las columnas correspondientes de la tabla de destino. El orden de la columna no es importante. Después de que se agregan los datos correctamente a la tabla de destino, la tabla de origen queda vacía. Por lo general, ALTER TABLE APPEND es mucho más rápida que una operación CREATE TABLE AS o INSERT INTO similar debido a que los datos se mueven, no se duplican.

> **Note**
>
> ALTER TABLE APPEND mueve bloques de datos entre la tabla de origen y la tabla de destino. Para mejorar el rendimiento, ALTER TABLE APPEND no compacta el almacenamiento como parte de la operación APPEND. Como resultado, el uso del almacenamiento aumenta temporalmente. Parar reclamar el espacio, ejecute una operación VACUUM.

Las columnas con los mismos nombres también deben tener atributos de columna idénticos. Si la tabla de origen o la tabla de destino contienen columnas que no existen en la otra tabla, use los parámetros IGNOREEXTRA o FILLTARGET para especificar cómo se deben administrar las columnas adicionales.

No se puede agregar una columna de identidad. El comando fallará si ambas tablas incluyen una columna de identidad. Si solo una tabla tiene una columna de identidad, incluya el parámetro FILLTARGET o IGNOREEXTRA. Para obtener más información, consulte Notas de uso acerca de ALTER TABLE APPEND.

Puede anexar una columna GENERATED BY DEFAULT AS IDENTITY. Puede actualizar las columnas definidas como GENERATED BY DEFAULT AS IDENTITY con los valores proporcionados. Para obtener más información, consulte Notas de uso acerca de ALTER TABLE APPEND.

La tabla de destino debe ser una tabla permanente. No obstante, el origen puede ser una tabla permanente o una vista materializada configurada para la ingesta de streaming. Ambos objetos deben usar el mismo estilo de distribución y la misma clave de distribución, si se ha definido una. Si los objetos están ordenados, ambos deben usar el mismo estilo de ordenación y definir las mismas columnas como claves de ordenación.

Un comando ALTER TABLE APPEND se confirma automáticamente apenas se completa la operación. No se puede revertir. No puede ejecutar ALTER TABLE APPEND en un bloque de transacción (BEGIN … END). Para obtener más información acerca de las transacciones, consulte Aislamiento serializable.

### Privilegios necesarios

Según el comando ALTER TABLE APPEND, se requiere uno de los siguientes privilegios:

- Superusuario
- Usuarios con el privilegio de sistema ALTER TABLE
- Usuarios con los privilegios DELETE y SELECT en la tabla de origen y el privilegio INSERT en la tabla de destino

## Sintaxis

```
ALTER TABLE target_table_name APPEND FROM [ source_table_name
 | source_materialized_view_name ]
[ IGNOREEXTRA | FILLTARGET ]
```

La adición desde una vista materializada solo funciona en el caso en que la vista materializada esté configurada para Ingesta de streaming a una vista materializada.

## Parámetros

target_table_name

El nombre de la tabla a la que se agregan las filas. Especifique solo el nombre de la tabla o use el formato schema_name.table_name (nombre-de_esquema.nombre_de_tabla) para usar un esquema específico. La tabla de destino debe ser una tabla permanente existente.

FROM source_table_name

El nombre de la tabla que proporciona las filas que se agregarán. Especifique solo el nombre de la tabla o use el formato schema_name.table_name (nombre-de_esquema.nombre_de_tabla) para usar un esquema específico. La tabla de origen debe ser una tabla permanente existente.

FROM source_materialized_view_name

El nombre de la vista materializada que proporciona las filas que se agregarán. La adición desde una vista materializada solo funciona en el caso en que la vista materializada esté configurada

para Ingesta de streaming a una vista materializada. La vista materializada de origen ya debe existir.

IGNOREEXTRA

Una palabra clave que especifica si la tabla de origen incluye columnas que no están presentes en la tabla de destino; los datos de las columnas adicionales deben descartarse. No puede utilizar IGNOREEXTRA con FILLTARGET.

FILLTARGET

Una palabra clave que especifica si la tabla de destino incluye columnas que no están presentes en la tabla de origen; las columnas deben completarse con el valor de la columna DEFAULT, si se definió uno, o con el valor NULL. No puede utilizar IGNOREEXTRA con FILLTARGET.

## Notas de uso acerca de ALTER TABLE APPEND

ALTER TABLE APPEND mueve solo las columnas idénticas de la tabla de origen a la tabla de destino. El orden de la columna no es importante.

Si la tabla de origen o la tabla de destino contienen columnas adicionales, utilice FILLTARGET o IGNOREEXTRA según las siguientes reglas:

- Si la tabla de origen contiene columnas que no existen en la tabla de destino, incluya IGNOREEXTRA. El comando ignora las columnas adicionales de la tabla de origen.
- Si la tabla de destino contiene columnas que no existen en la tabla de origen, incluya FILLTARGET. El comando rellena las columnas adicionales de la tabla de destino con el valor de columna predeterminado, con el valor IDENTITY, si se definió uno, o con NULL.
- El comando fallará si ambas tablas, la de origen y la de destino, contienen columnas adicionales. No pueden utilizar ambos parámetros, FILLTARGET e IGNOREEXTRA.

El comando fallará si hay una columna con el mismo nombre, pero con diferentes atributos, en ambas tablas. Las columnas con nombres similares deben tener los siguientes atributos en común:

- Tipo de datos:
- Tamaño de la columna
- Codificación de compresión
- Sin valores nulos
- Estilo de ordenación

- Columnas con clave de ordenación
- Estilo de distribución
- Columnas con clave de distribución

No se puede agregar una columna de identidad. El comando fallará si ambas tablas, la de origen y la de destino, tienen columnas de identidad. Si solo la tabla de origen tiene una columna de identidad, incluya el parámetro IGNOREEXTRA para ignorar la columna de identidad. Si solo la tabla de destino tiene una columna de identidad, incluya el parámetro FILLTARGET para completar la columna de identidad según la cláusula IDENTITY definida para la tabla. Para obtener más información, consulte DEFAULT.

Puede anexar una columna de identidad predeterminada con la instrucción ALTER TABLE APPEND. Para obtener más información, consulte CREATE TABLE.

## Ejemplos de ALTER TABLE APPEND

Supongamos que su organización mantiene una tabla, SALES_MONTHLY, para capturar transacciones de venta actuales. Desea mover datos de la tabla de transacción a la tabla SALES, todos los meses.

Puede utilizar los comandos INSERT INTO y TRUNCATE para llevar a cabo la tarea.

```
insert into sales (select * from sales_monthly);
truncate sales_monthly;
```

No obstante, puede realizar la misma operación de una manera mucho más eficiente con el comando ALTER TABLE APPEND.

Primero, consulte la tabla de catálogo del sistema PG_TABLE_DEF para verificar que ambas tablas tengan las mismas columnas con atributos de columna idénticos.

```
select trim(tablename) as table, "column", trim(type) as type,
encoding, distkey, sortkey, "notnull"
from pg_table_def where tablename like 'sales%';

table      | column     | type                          | encoding | distkey | sortkey |
 notnull
-----------+------------+-------------------------------+----------+---------+---------
+--------
```

```
sales      | salesid    | integer                     | lzo      | false   |       0 |
 true
sales      | listid     | integer                     | none     | true    |       1 |
 true
sales      | sellerid   | integer                     | none     | false   |       2 |
 true
sales      | buyerid    | integer                     | lzo      | false   |       0 |
 true
sales      | eventid    | integer                     | mostly16 | false   |       0 |
 true
sales      | dateid     | smallint                    | lzo      | false   |       0 |
 true
sales      | qtysold    | smallint                    | mostly8  | false   |       0 |
 true
sales      | pricepaid  | numeric(8,2)                | delta32k | false   |       0 |
 false
sales      | commission | numeric(8,2)                | delta32k | false   |       0 |
 false
sales      | saletime   | timestamp without time zone | lzo      | false   |       0 |
 false
salesmonth | salesid    | integer                     | lzo      | false   |       0 |
 true
salesmonth | listid     | integer                     | none     | true    |       1 |
 true
salesmonth | sellerid   | integer                     | none     | false   |       2 |
 true
salesmonth | buyerid    | integer                     | lzo      | false   |       0 |
 true
salesmonth | eventid    | integer                     | mostly16 | false   |       0 |
 true
salesmonth | dateid     | smallint                    | lzo      | false   |       0 |
 true
salesmonth | qtysold    | smallint                    | mostly8  | false   |       0 |
 true
salesmonth | pricepaid  | numeric(8,2)                | delta32k | false   |       0 |
 false
salesmonth | commission | numeric(8,2)                | delta32k | false   |       0 |
 false
salesmonth | saletime   | timestamp without time zone | lzo      | false   |       0 |
 false
```

A continuación, observe el tamaño de cada tabla.

```
select count(*) from sales_monthly;
 count
-------
  2000
(1 row)

select count(*) from sales;
 count
-------
 412,214
(1 row)
```

Ahora ejecute el siguiente comando ALTER TABLE APPEND.

```
alter table sales append from sales_monthly;
```

Observe el tamaño de cada tabla nuevamente. La tabla SALES_MONTHLY ahora tiene 0 filas, y la tabla SALES tiene 2 000 filas más.

```
select count(*) from sales_monthly;
 count
-------
     0
(1 row)

select count(*) from sales;
 count
-------
 414214
(1 row)
```

Si la tabla de origen contiene más columnas que la tabla de destino, especifique el parámetro IGNOREEXTRA. En el siguiente ejemplo, se usa el parámetro IGNOREEXTRA para ignorar las columnas adicionales de la tabla SALES_LISTING cuando agrega columnas en la tabla SALES.

```
alter table sales append from sales_listing ignoreextra;
```

Si la tabla de destino contiene más columnas que la tabla de origen, especifique el parámetro FILLTARGET. En el siguiente ejemplo, se usa el parámetro FILLTARGET para completar columnas en la tabla SALES_REPORT que no existen en la tabla SALES_MONTH.

```
alter table sales_report append from sales_month filltarget;
```

A continuación, se muestra un ejemplo de cómo utilizar ALTER TABLE APPEND con una vista materializada como origen.

```
ALTER TABLE target_tbl APPEND FROM my_streaming_materialized_view;
```

Los nombres de la tabla y de la vista materializada de este ejemplo son muestras. La adición desde una vista materializada solo funciona en el caso en que la vista materializada esté configurada para Ingesta de streaming a una vista materializada. Mueve todos los registros de la vista materializada de origen a una tabla de destino con el mismo esquema que la vista materializada y deja la vista materializada intacta. Este es el mismo comportamiento que cuando el origen de los datos es una tabla.

# ALTER USER

Cambia un usuario de base de datos.

## Privilegios necesarios

Los siguientes privilegios son necesarios para ALTER USER:

- Superusuario
- Usuarios con el privilegio ALTER USER
- Usuario actual que desea cambiar su propia contraseña

## Sintaxis

```
ALTER USER username [ WITH ] option [, ... ]

where option is

CREATEDB | NOCREATEDB
| CREATEUSER | NOCREATEUSER
| SYSLOG ACCESS { RESTRICTED | UNRESTRICTED }
| PASSWORD { 'password' | 'md5hash' | 'sha256hash' | DISABLE }
[ VALID UNTIL 'expiration_date' ]
| RENAME TO new_name |
```

```
| CONNECTION LIMIT { limit | UNLIMITED }
| SESSION TIMEOUT limit | RESET SESSION TIMEOUT
| SET parameter { TO | = } { value | DEFAULT }
| RESET parameter
| EXTERNALID external_id
```

## Parámetros

username

Nombre del usuario.

WITH

Palabra clave opcional.

CREATEDB | NOCREATEDB

La opción CREATEDB permite que el usuario cree nuevas bases de datos. NOCREATEDB es el valor predeterminado.

CREATEUSER | NOCREATEUSER

La opción CREATEUSER crea un superusuario con todos los privilegios de base de datos, incluido CREATE USER. El valor predeterminado es NOCREATEUSER. Para obtener más información, consulte superuser.

SYSLOG ACCESS { RESTRICTED | UNRESTRICTED }

Se trata de una cláusula que especifica el nivel de acceso a las vistas y las tablas de sistema de Amazon Redshift que tiene el usuario.

Los usuarios normales que tengan el permiso SYSLOG ACCESS RESTRICTED solo pueden ver las filas generadas por ese usuario en las tablas y vistas del sistema visibles para el usuario. El valor predeterminado es RESTRICTED.

Los usuarios normales que tengan el permiso SYSLOG ACCESS UNRESTRICTED pueden ver todas las filas en las tablas y vistas del sistema visibles para el usuario, incluidas las filas generadas por otro usuario. UNRESTRICTED no permite que los usuarios normales puedan obtener acceso a las tablas visibles para los superusuarios. Solo los superusuarios pueden ver estas tablas.

> **Note**
>
> Al brindar a un usuario acceso sin restricciones a las tablas del sistema, le proporciona la visibilidad necesaria para ver los datos generados por otros usuarios. Por ejemplo, STL_QUERY y STL_QUERYTEXT contienen todo el texto de las instrucciones INSERT, UPDATE y DELETE, que podrían incluir datos confidenciales generados por los usuarios.

Todas las filas de SVV_TRANSACTIONS son visibles para todos los usuarios.

Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

PASSWORD { 'password' | 'md5hash' | 'sha256hash' | DISABLE }

Establece la contraseña del usuario.

De manera predeterminada, los usuarios pueden cambiar sus propias contraseñas, a menos que la contraseña esté deshabilitada. Para deshabilitar la contraseña de un usuario, especifique DISABLE. Cuando se deshabilita la contraseña de un usuario, se elimina del sistema y el usuario solo puede iniciar sesión con credenciales de usuario de AWS Identity and Access Management IAM temporales. Para obtener más información, consulte Uso de la autenticación de IAM para generar credenciales de usuario de base de datos. Solo un superusuario puede habilitar o deshabilitar contraseñas. No puede deshabilitar la contraseña de un superusuario. Para habilitar una contraseña, ejecute ALTER USER y especifique una contraseña.

Para obtener información acerca de cómo utilizar el parámetro PASSWORD, consulte CREAR USUARIO.

VALID UNTIL 'expiration_date'

Especifica que la contraseña tiene una fecha de vencimiento. Use el valor `'infinity'` para evitar tener una fecha de vencimiento. El tipo de datos válido para este parámetro es timestamp (marca temporal).

Solo los superusuarios pueden utilizar este parámetro.

RENAME TO

Cambia el nombre del usuario.

new_name

Nuevo nombre de usuario. Para obtener más información acerca de los nombres válidos, consulte Nombres e identificadores.

> **Important**
>
> Cuando cambia el nombre de un usuario, también debe restablecer la contraseña del usuario. La contraseña de restablecimiento no tiene por qué ser diferente de la contraseña anterior. Debido a que el nombre del usuario se utiliza como parte del cifrado de la contraseña, la contraseña se borra cada vez que se cambia el nombre de un usuario. El usuario no podrá iniciar sesión hasta que no se haya restablecido la contraseña. Por ejemplo:
>
> ```
> alter user newuser password 'EXAMPLENewPassword11';
> ```

CONNECTION LIMIT { limit | UNLIMITED }

La cantidad máxima de conexiones a la base de datos que el usuario puede tener abiertas al mismo tiempo. Este límite no se aplica a los superusuarios. Use la palabra clave UNLIMITED para permitir la cantidad máxima de conexiones simultáneas. También puede aplicar un límite en la cantidad de conexiones de cada base de datos. Para obtener más información, consulte CREATE DATABASE. El valor predeterminado es UNLIMITED. Para ver las conexiones actuales, consulte la vista del sistema STV_SESSIONS.

> **Note**
>
> Si se aplican los límites de conexión tanto para usuarios como para bases de datos, debe haber una ranura de conexión sin utilizar disponible dentro de ambos límites cuando un usuario intenta conectarse.

SESSION TIMEOUT limit | RESET SESSION TIMEOUT

El tiempo máximo en segundos durante el cual una sesión permanece inactiva o en reposo. El rango es de 60 segundos (un minuto) a 1 728 000 segundos (20 días). Si no se establece el tiempo de espera de la sesión para el usuario, se aplica el ajuste del clúster. Para obtener más información, consulte Cuotas y límites de Amazon Redshift en la Guía de administración de Amazon Redshift.

Cuando se establece el tiempo de espera de la sesión, se aplica solo a las sesiones nuevas.

Para ver información sobre las sesiones de usuario activas, incluidos la hora de inicio, el nombre de usuario y el tiempo de espera de la sesión, consulte la vista de sistema STV_SESSIONS. Para ver información sobre el historial de sesiones del usuario, consulte la vista STL_SESSIONS. Para recuperar información acerca de los usuarios de las bases de datos, incluidos los valores de tiempo de espera de la sesión, consulte la vista SVL_USER_INFO.

SET

Establece un parámetro de configuración a un nuevo valor predeterminado para todas las sesiones ejecutadas por el usuario especificado.

RESET

Restablece un parámetro de configuración al valor predeterminado original para el usuario especificado.

parameter

Nombre del parámetro que se debe establecer o restablecer.

value

Valor nuevo del parámetro.

DEFAULT

Establece el parámetro de configuración al valor predeterminado para todas las sesiones ejecutadas por el usuario especificado.

EXTERNALID external_id

El identificador del usuario, que está asociado a un proveedor de identidades. El usuario debe tener desactivada la contraseña. Para obtener más información, consulte Federación de proveedores de identidades (IdP) nativos para Amazon Redshift.

## Notas de uso

- Intento de modificar rdsdb: no puede modificar al usuario denominado `rdsdb`.
- Creación de una contraseña desconocida: cuando use la autenticación de AWS Identity and Access Management (IAM) para crear credenciales de usuario de la base de datos, es posible que quiera crear un superusuario que solo pueda iniciar sesión con credenciales temporales. Puede deshabilitar la contraseña de un superusuario, pero no puede crear una contraseña desconocida mediante una cadena hash MD5 generada aleatoriamente.

```
alter user iam_superuser password 'md512345678901234567801234567890l2';
```

- Configuración de search_path: cuando establece el parámetro search_path con el comando ALTER USER, la modificación entra en vigor en el próximo inicio de sesión del usuario especificado. Si desea cambiar el valor search_path para la sesión y el usuario actuales, utilice un comando SET.
- Configuración de la zona horaria: cuando utilice ESTABLECER ZONA HORARIA con el comando ALTER USER, la modificación entra en vigor en el próximo inicio de sesión del usuario especificado.
- Trabajo con políticas de seguridad a nivel de fila y enmascaramiento de datos dinámico: cuando el clúster o espacio de nombres sin servidor aprovisionado tiene políticas de enmascaramiento de datos dinámico o de seguridad en el nivel de fila, los siguientes comandos están bloqueados para los usuarios habituales:

  ```
  ALTER <current_user> SET enable_case_sensitive_super_attribute/
  enable_case_sensitive_identifier/downcase_delimited_identifier
  ```

  Solo los superusuarios y los usuarios con el privilegio ALTER USER pueden establecer estas opciones de configuración. Para obtener información sobre seguridad de nivel de fila, consulte Seguridad de nivel básico. Para obtener información sobre el enmascaramiento dinámico de datos, consulte Enmascaramiento de datos dinámico.

## Ejemplos

En el siguiente ejemplo, se le otorga al usuario ADMIN el privilegio de crear bases de datos:

```
alter user admin createdb;
```

En el siguiente ejemplo, se establece la contraseña del usuario ADMIN a `adminPass9` y se establece una fecha y hora de vencimiento para la contraseña:

```
alter user admin password 'adminPass9'
valid until '2017-12-31 23:59';
```

En el siguiente ejemplo, se cambia el nombre del usuario ADMIN a SYSADMIN:

```
alter user admin rename to sysadmin;
```

En el siguiente ejemplo, se actualiza el tiempo de espera de la sesión inactiva para un usuario a 300 segundos.

```
ALTER USER dbuser SESSION TIMEOUT 300;
```

Restablece el tiempo de espera de la sesión inactiva del usuario. Cuando lo restablece, se aplica el ajuste del clúster. Debe ser un superusuario de base de datos para ejecutar este comando. Para obtener más información, consulte Cuotas y límites de Amazon Redshift en la Guía de administración de Amazon Redshift.

```
ALTER USER dbuser RESET SESSION TIMEOUT;
```

En el ejemplo siguiente se actualiza el ID externo de un usuario denominado bob. El espacio de nombres es `myco_aad`. Si el espacio de nombres no está asociado a un proveedor de identidades registrado, se produce un error.

```
ALTER USER myco_aad:bob EXTERNALID "ABC123" PASSWORD DISABLE;
```

En el siguiente ejemplo se establece la zona horaria para todas las sesiones ejecutadas por un usuario específico de la base de datos. Cambia la zona horaria para las sesiones posteriores, pero no para la actual.

```
ALTER USER odie SET TIMEZONE TO 'Europe/Zurich';
```

En el siguiente ejemplo, se establece el número máximo de conexiones a bases de datos que el usuario bob puede tener abiertas.

```
ALTER USER bob CONNECTION LIMIT 10;
```

# ANALYZE

Actualiza las estadísticas de la tabla para que el planificador de consultas las utilice.

## Privilegios necesarios

Los siguientes privilegios son necesarios para ANALYZE:

- Superusuario
- Usuarios con el privilegio ANALYZE
- Propietario de la relación
- Propietario de la base de datos con quien se comparte la tabla

## Sintaxis

```
ANALYZE [ VERBOSE ]
[ [ table_name [ ( column_name [, ...] ) ] ]
[ PREDICATE COLUMNS | ALL  COLUMNS ]
```

## Parámetros

VERBOSE

Una cláusula que devuelve mensajes de información de progreso acerca de la operación ANALYZE. Esta opción es útil cuando no se especifica una tabla.

table_name

Puede analizar tablas específicas, incluidas tablas temporales. Puede completar la tabla con el nombre de su esquema. Si lo desea, puede especificar un table_name (nombre_de_tabla) para analizar una única tabla. No puede especificar más de un table_name (nombre_de_tabla) con una única instrucción ANALYZE table_name (nombre_de_tabla). Si no se especifica un valor table_name, se analizan todas las tablas de la base de datos actualmente conectada, incluidas las tablas persistentes en el catálogo del sistema. Amazon Redshift omite el análisis de una tabla si el porcentaje de filas que se han modificado desde la última operación ANALYZE es inferior al límite de análisis. Para obtener más información, consulte Umbral de análisis.

No necesita analizar las tablas de sistema de Amazon Redshift (tablas STL y STV).

column_name

Si especifica un valor table_name (nombre_de_tabla), también puede especificar una o varias columnas de la tabla (como una lista de valores separados por columnas entre paréntesis). Si se especifica una lista de columnas, solo se analizarán las columnas que aparecen en la lista.

PREDICATE COLUMNS | ALL COLUMNS

Se trata de cláusulas que indican si ANALYZE debe incluir solo columnas de predicados. Especifique PREDICATE COLUMNS para analizar solamente las columnas utilizadas

como predicados en consultas anteriores o que sean posibles candidatas para utilizarse como predicados. Especifique ALL COLUMNS para analizar todas las columnas. El valor predeterminado es ALL COLUMNS.

Se incluye una columna en el conjunto de columnas de predicados si cualquiera de los siguientes es verdadero:

- La columna ha sido utilizada en una consulta como parte de un filtro, una condición de combinación o una cláusula de agrupación.
- La columna es una clave de distribución.
- La columna forma parte de una clave de ordenación.

Si no se marca ninguna columna como columna de predicado, por ejemplo porque la tabla aún no se ha consultado, se analizan todas las columnas incluso si PREDICATE COLUMNS está especificado. Cuando esto sucede, Amazon Redshift podría responder con un mensaje similar a No se han encontrado columnas de predicado para "*nombre-de-tabla*". Analizando todas las columnas. Para obtener más información acerca de las columnas de predicados, consulte Análisis de tablas.

## Notas de uso

Amazon Redshift ejecuta ANALYZE de manera automática en las tablas que usted crea con los siguientes comandos:

- CREATE TABLE AS
- CREATE TEMP TABLE AS
- SELECT INTO

No puede analizar una tabla externa.

No es necesario ejecutar el comando ANALYZE en estas tablas cuando se crean por primera vez. Si las modifica, debe analizarlas de la misma manera en que analiza las demás tablas.

Umbral de análisis

Para reducir el tiempo de procesamiento y mejorar el rendimiento general del sistema, Amazon Redshift omite la operación ANALYZE en una tabla si el porcentaje de filas que se han modificado desde que se ejecutó el comando ANALYZE por última vez es inferior

al límite de análisis especificado por el parámetro analyze_threshold_percent. Por defecto, analyze_threshold_percent es 10. Si desea cambiar analyze_threshold_percent para la sesión actual, ejecute el comando SET. En el siguiente ejemplo, se cambia analyze_threshold_percent a 20 por ciento.

```
set analyze_threshold_percent to 20;
```

Para analizar las tablas cuando solo han cambiado unas pocas filas, establezca analyze_threshold_percent en un número pequeño elegido arbitrariamente. Por ejemplo, si establece analyze_threshold_percent como 0,01, las tablas con 100 000 000 filas no se omitirán si se han modificado al menos 10 000 filas.

```
set analyze_threshold_percent to 0.01;
```

Si ANALYZE omite una tabla porque no cumple con el umbral de análisis, Amazon Redshift devuelve el siguiente mensaje.

```
ANALYZE SKIP
```

Para analizar todas las tablas, aunque no haya cambiado ninguna fila, establezca analyze_threshold_percent en 0.

Para ver los resultados de las operaciones ANALYZE, consulte la tabla del sistema STL_ANALYZE.

Para obtener más información acerca del análisis de tablas, consulte Análisis de tablas.

## Ejemplos

Analiza todas las tablas de la base de datos TICKIT y devolver la información de progreso.

```
analyze verbose;
```

Analiza solo la tabla LISTING.

```
analyze listing;
```

Analiza las columnas VENUEID y VENUENAME de la tabla VENUE.

```
analyze venue(venueid, venuename);
```

Analiza solo columnas de predicado de la tabla VENUE.

```
analyze venue predicate columns;
```

# ANALYZE COMPRESSION

Realiza un análisis de compresión y produce un informe con la codificación de compresión sugerida para las tablas analizadas. Para cada columna, el informe incluye un cálculo de la reducción potencial de espacio en disco en comparación con la codificación RAW.

## Sintaxis

```
ANALYZE COMPRESSION
[ [ table_name ]
[ ( column_name [, ...] ) ] ]
[COMPROWS numrows]
```

## Parámetros

table_name

Puede analizar la compresión para tablas específicas, incluidas tablas temporales. Puede completar la tabla con el nombre de su esquema. Si lo desea, puede especificar un table_name (nombre_de_tabla) para analizar una única tabla. Si no especifica table_name (nombre_de_tabla), se analizarán todas las tablas de la base de datos que está conectada actualmente. No puede especificar varios valores table_name (nombre_de_tabla) con una única instrucción ANALYZE COMPRESSION.

column_name

Si especifica un valor table_name (nombre_de_tabla), también puede especificar una o varias columnas de la tabla (como una lista de valores separados por columnas entre paréntesis).

COMPROWS

Cantidad de filas que se usarán como el tamaño de muestra para el análisis de compresión. El análisis se ejecuta sobre las filas de cada sector de datos. Por ejemplo, si especifica COMPROWS 1000000 (1 000 000) y el sistema contiene 4 sectores totales, no se leen y analizan más de 250 000 filas por sector. Si no especifica COMPROWS, el tamaño de muestra se establecerá de manera predeterminada en 100 000 por sector. Los valores de COMPROWS inferiores al valor predeterminado de 100 000 filas por sector se actualizan automáticamente

al valor predeterminado. No obstante, el análisis de compresión no produce recomendaciones si la cantidad de datos de la tabla no alcanza para producir una muestra significativa. Si el número de COMPROWS es superior a la cantidad de filas de la tabla, el comando ANALYZE COMPRESSION avanza y ejecuta el análisis de compresión contra todas las filas disponibles El uso de COMPROWS produce un error si no se especifica una tabla.

numrows

Cantidad de filas que se usarán como el tamaño de muestra para el análisis de compresión. El rango aceptado para numrows (número_de_filas) es un número comprendido entre 1000 y 1000000000 (1 000 000 000).

## Notas de uso

ANALYZE COMPRESSION adquiere un bloqueo de tabla exclusivo, que previene lecturas y escrituras concurrentes en la tabla. Ejecute solo el comando ANALYZE COMPRESSION cuando la tabla esté inactiva.

Ejecute ANALYZE COMPRESSION para obtener recomendaciones para esquemas de codificación de columna, según una muestra del contenido de la tabla. ANALYZE COMPRESSION es una herramienta de consultoría y no modifica las codificaciones de columna de la tabla. Puede aplicar la codificación sugerida creando de nuevo la tabla o generando una nueva tabla con el mismo esquema. El espacio en disco se puede reducir significativamente creando de nuevo una tabla sin comprimir con los esquemas de codificación adecuados. Este enfoque ahorra espacio en disco y mejora el rendimiento de las consultas en las cargas de trabajo relacionadas con operaciones de E/S.

ANALYZE COMPRESSION omite la verdadera fase de análisis y devuelve directamente el tipo de codificación original de cualquier columna que se haya designado como SORTKEY. Esto se debe a que los análisis restringidos por rango pueden funcionar mal cuando las columnas SORTKEY están mucho más comprimidas que otras columnas.

## Ejemplos

En el siguiente ejemplo, se muestran la codificación y la reducción porcentual estimada para las columnas de la tabla LISTING únicamente:

```
analyze compression listing;
```

```
  Table  |     Column     | Encoding | Est_reduction_pct
---------+----------------+----------+-------------------
 listing | listid         | az64     | 40.96
 listing | sellerid       | az64     | 46.92
 listing | eventid        | az64     | 53.37
 listing | dateid         | raw      | 0.00
 listing | numtickets     | az64     | 65.66
 listing | priceperticket | az64     | 72.94
 listing | totalprice     | az64     | 68.05
 listing | listtime       | az64     | 49.74
```

En el siguiente ejemplo, se analizan las columnas QTYSOLD, COMMISSION y SALETIME de la tabla SALES.

```
analyze compression sales(qtysold, commission, saletime);

 Table |   Column   | Encoding | Est_reduction_pct
-------+------------+----------+-------------------
 sales | salesid    | N/A      | 0.00
 sales | listid     | N/A      | 0.00
 sales | sellerid   | N/A      | 0.00
 sales | buyerid    | N/A      | 0.00
 sales | eventid    | N/A      | 0.00
 sales | dateid     | N/A      | 0.00
 sales | qtysold    | az64     | 83.06
 sales | pricepaid  | N/A      | 0.00
 sales | commission | az64     | 71.85
 sales | saletime   | az64     | 49.63
```

# ATTACH MASKING POLICY

Adjunta una política de enmascaramiento dinámico de datos existente a una columna. Para obtener más información sobre el enmascaramiento dinámico de datos, consulte Enmascaramiento de datos dinámico.

Los superusuarios y los usuarios o roles que tengan el rol sys:secadmin pueden adjuntar una política de enmascaramiento.

## Sintaxis

```
ATTACH MASKING POLICY policy_name
```

```
  ON { relation_name }
  ( {output_columns_names | output_path} ) [ USING ( {input_column_names | input_path
)} ]
  TO { user_name | ROLE role_name | PUBLIC }
  [ PRIORITY priority ];
```

## Parámetros

policy_name

El nombre de la política de enmascaramiento que se va a adjuntar.

relation_name

Es el nombre de la relación a la que se asociará la política de enmascaramiento.

output_column_names

Los nombres de las columnas a las que se aplicará la política de enmascaramiento.

output_paths

Es la ruta completa del objeto SUPER al que se aplicará la política de enmascaramiento, incluido el nombre de la columna. Por ejemplo, para una relación con una columna de tipo SUPER llamada `person`, output_path podría ser `person.name.first_name`.

input_column_names

Los nombres de las columnas que la política de enmascaramiento tomará como entrada. Este parámetro es opcional. Si no se especifica, la política de enmascaramiento utiliza output_column_names como entradas.

input_paths

Es la ruta completa del objeto SUPER que la política de enmascaramiento tomará como entrada. Este parámetro es opcional. Si no se especifica, la política de enmascaramiento utiliza output_path como entradas.

user_name

El nombre del usuario al que se adjuntará la política de enmascaramiento. No puede adjuntar dos políticas a la misma combinación de usuario y columna o rol y columna. Puede adjuntar una política a un usuario y otra al rol del usuario. En este caso, se aplica la política de mayor prioridad.

Solo puede establecer user_name, role_name o PUBLIC en un solo comando ATTACH MASKING POLICY.

role_name

El nombre del rol al que se adjuntará la política de enmascaramiento. No puede adjuntar dos políticas al mismo par de columna-rol. Puede adjuntar una política a un usuario y otra política al rol del usuario. En este caso, se aplica la política de mayor prioridad.

Solo puede establecer user_name, role_name o PUBLIC en un solo comando ATTACH MASKING POLICY.

PUBLIC

Adjunta la política de enmascaramiento a todos los usuarios que acceden a la tabla. Para que se apliquen, debe dar a otras políticas de enmascaramiento adjuntas a pares de columna-usuario o columna-rol específicos una prioridad mayor que a la política PUBLIC.

Solo puede establecer user_name, role_name o PUBLIC en un solo comando ATTACH MASKING POLICY.

priority

Prioridad de la política de enmascaramiento. Cuando se aplican varias políticas de enmascaramiento a la consulta de un usuario determinado, se aplica la de mayor prioridad.

No puede asociar dos políticas diferentes a la misma columna con la misma prioridad, incluso si las dos políticas están asociadas a usuarios o roles diferentes. Puede asociar la misma política varias veces al mismo conjunto de parámetros de tabla, columna de salida, columna de entrada y prioridad, siempre que el usuario o rol al que se asocie la política sea diferente cada vez.

No puede aplicar una política a una columna con la misma prioridad que otra política adjunta a esa columna, aunque sean para roles diferentes. Este campo es opcional. Si no especifica una prioridad, la política de enmascaramiento establece de forma predeterminada una prioridad de 0.

# ATTACH RLS POLICY

Adjunte una política de seguridad de la fila en una tabla a uno o más usuarios o roles.

Los superusuarios y los usuarios o roles que tienen el rol `sys:secadmin` puede adjuntar una política.

## Sintaxis

```
ATTACH RLS POLICY policy_name ON [TABLE] table_name [, ...]
TO { user_name | ROLE role_name | PUBLIC } [, ...]
```

## Parámetros

policy_name

El nombre de la política de .

ON [TABLE] table_name [, ...]

Es la relación a la que se asocia la política de seguridad de nivel de fila.

TO { user_name | ROLE role_name | PUBLIC} [, ...]

Especifica si la política se adjunta a uno o varios usuarios o roles especificados.

## Notas de uso

Al trabajar con la instrucción ATTACH RLS POLICY, observe lo siguiente:

- La tabla que se adjunta debe tener todas las columnas enumeradas en la cláusula WITH de la instrucción de creación de políticas.
- Amazon Redshift RLS no permite asociar políticas de RLS a los siguientes objetos:
  - Tablas
  - Vistas
  - Vistas de enlace en tiempo de ejecución
  - Vistas materializadas
- Amazon Redshift RLS no permite asociar políticas de RLS a los siguientes objetos:
  - Tablas de catálogo
  - Relaciones entre bases de datos
  - tablas externas
  - Tablas temporales
  - Tablas de búsqueda de políticas
  - Tablas básicas de vistas materializadas

- Las políticas de RLS que se asocian a los superusuarios o a los usuarios con el permiso `sys:secadmin` se omiten.

### Ejemplos

En el siguiente ejemplo, se adjunta una política en una tabla a un rol.

```
ATTACH RLS POLICY policy_concerts ON tickit_category_redshift TO ROLE analyst, ROLE
 dbadmin;
```

## BEGIN

Inicia una transacción. Sinónimo de START TRANSACTION.

Una transacción es una única unidad lógica de trabajo, independientemente de tener un comando o varios comandos. Por lo general, todos los comandos de una transacción se ejecutan en una instantánea de la base de datos cuya hora de inicio se determina con el valor establecido para el parámetro de configuración del sistema `transaction_snapshot_begin`.

De manera predeterminada, las operaciones individuales de Amazon Redshift (consultas, instrucciones DDL, cargas) se confirman de manera automática en la base de datos. Si desea suspender la confirmación de una operación hasta que se complete el trabajo subsiguiente, necesita abrir una transacción con la instrucción BEGIN, luego ejecutar los comandos requeridos y cerrar la transacción con una instrucción COMMIT o END. Si es necesario, puede utilizar una instrucción ROLLBACK para detener una transacción en curso. Una excepción a este comportamiento es el comando TRUNCATE, que confirma la transacción en la que se ejecuta y no puede revertirse.

### Sintaxis

```
BEGIN [ WORK | TRANSACTION ] [ ISOLATION LEVEL option ] [ READ WRITE | READ ONLY ]

START TRANSACTION [ ISOLATION LEVEL option ] [ READ WRITE | READ ONLY ]

Where option is

SERIALIZABLE
| READ UNCOMMITTED
| READ COMMITTED
| REPEATABLE READ
```

```
Note: READ UNCOMMITTED, READ COMMITTED, and REPEATABLE READ have no
operational impact and map to SERIALIZABLE in Amazon Redshift. You can see database
 isolation levels on your cluster
by querying the stv_db_isolation_level table.
```

## Parámetros

WORK

Palabra clave opcional.

TRANSACTION

Palabra clave opcional; WORK y TRANSACTION son sinónimos.

ISOLATION LEVEL SERIALIZABLE

El aislamiento serializable es compatible de manera predeterminada, por lo que el comportamiento de la transacción es el mismo si la sintaxis está incluida o no en la instrucción. Para obtener más información, consulte Administración de operaciones de escritura simultáneas. No se admite otro nivel de aislamiento.

> **Note**
>
> El estándar de SQL define cuatro niveles de aislamiento de transacciones para prevenir lecturas sucias (donde una transacción lee los datos escritos por una transacción paralela sin confirmar), lecturas no repetibles (donde una transacción relee datos que leyó anteriormente y descubre que otra transacción que se confirmó después de la lectura inicial modificó los datos) y lecturas fantasmas (donde una transacción vuelve a ejecutar una consulta, devuelve un conjunto de filas que cumplen una condición de búsqueda y luego descubre que el conjunto de filas ha cambiado debido a otra transacción recientemente confirmada):
>
> - Lectura sin confirmar: las lecturas sucias, lecturas no repetibles y lecturas fantasmas son posibles.
> - Lectura confirmada: las lecturas no repetibles y lecturas fantasmas son posibles.
> - Lectura repetible: las lecturas fantasmas son posibles.
> - Serializable: previene lecturas sucias, lecturas no repetibles y lecturas fantasmas.

Aunque puede utilizar cualquiera de los cuatro niveles de aislamiento de las transacciones, Amazon Redshift procesa todos los niveles de aislamiento como serializables.

READ WRITE

Le otorga a la transacción permisos de lectura y escritura.

READ ONLY

Le otorga a la transacción permisos de solo lectura.

### Ejemplos

En el siguiente ejemplo, se inicia un bloque de transacción serializable:

```
begin;
```

En el siguiente ejemplo, se inicia el bloque de transacción con un nivel de aislamiento serializable y permisos de lectura y escritura:

```
begin read write;
```

## CALL

Ejecuta un procedimiento almacenado. El comando CALL debe incluir el nombre del procedimiento y los valores del argumento de entrada. Debe llamar a un procedimiento almacenado mediante la instrucción CALL.

**Note**

CALL no puede formar parte de ninguna consultar normal.

### Sintaxis

```
CALL sp_name ( [ argument ] [, ...] )
```

## Parámetros

sp_name

El nombre del procedimiento que debe ejecutarse.

argument

El valor del argumento de entrada. Este parámetro también puede ser un nombre de función; por ejemplo `pg_last_query_id()`. No es posible utilizar consultas como argumentos de CALL.

## Notas de uso

Los procedimientos almacenados de Amazon Redshift admiten llamadas anidadas y recursivas, tal y como se describe a continuación. Además, asegúrese de que el soporte de los controladores está al día, también se describe a continuación.

Temas

- Llamadas anidadas
- Soporte de los controladores

### Llamadas anidadas

Los procedimientos almacenados de Amazon Redshift admiten llamadas anidadas y recursivas. El número máximo de niveles de agrupación permitido es 16. Las llamadas agrupadas pueden encapsular la lógica empresarial en procedimientos más pequeños que pueden compartir varios remitentes.

Si llama a un procedimiento agrupado que tiene parámetros de salida, el procedimiento interior debe definir argumentos INOUT. En este caso, el procedimiento interior se pasa en una variable no constante. No se permiten argumentos OUT. Este comportamiento se produce porque se necesita una variable para retener la salida de la llamada interior.

La relación entre los procedimientos interior y exterior se registra en la columna `from_sp_call` de SVL_STORED_PROC_CALL.

El siguiente ejemplo muestra el paso de variable a una llamada de procedimientos agrupada a través de argumentos INOUT.

```
CREATE OR REPLACE PROCEDURE inner_proc(INOUT a int, b int, INOUT c int) LANGUAGE
 plpgsql
AS $$
BEGIN
  a := b * a;
  c := b * c;
END;
$$;

CREATE OR REPLACE PROCEDURE outer_proc(multiplier int) LANGUAGE plpgsql
AS $$
DECLARE
  x int := 3;
  y int := 4;
BEGIN
  DROP TABLE IF EXISTS test_tbl;
  CREATE TEMP TABLE test_tbl(a int, b varchar(256));
  CALL inner_proc(x, multiplier, y);
  insert into test_tbl values (x, y::varchar);
END;
$$;

CALL outer_proc(5);

SELECT * from test_tbl;
 a  | b
----+----
 15 | 20
(1 row)
```

Soporte de los controladores

Recomendamos actualizar los controladores Java Database Connectivity (JDBC) y Open Database Connectivity (ODBC) a la versión más reciente compatible con los procedimientos almacenados de Amazon Redshift.

Es posible que pueda utilizar el controlador existente si la herramienta de su cliente utiliza operaciones de la API del controlador que pasa a través de la instrucción CALL al servidor. Los parámetros de salida, si los hubiere, se devuelven como un conjunto de resultados de una fila.

Las versiones más recientes de los controladores de JDBC y ODBC de Amazon Redshift admiten los metadatos para el descubrimiento de procedimientos almacenados. También tienen soporte

`CallableStatement` para aplicaciones Java personalizadas. Para obtener más información acerca de los controladores, consulte Conexión a un clúster de Amazon Redshift mediante herramientas de cliente SQL en la Guía de administración de Amazon Redshift.

Los siguientes ejemplos muestran cómo usar diferentes operaciones de la API del controlador JDBC para llamadas de procedimientos almacenadas.

```
void statement_example(Connection conn) throws SQLException {
  statement.execute("CALL sp_statement_example(1)");
}

void prepared_statement_example(Connection conn) throws SQLException {
  String sql = "CALL sp_prepared_statement_example(42, 84)";
  PreparedStatement pstmt = conn.prepareStatement(sql);
  pstmt.execute();
}

void callable_statement_example(Connection conn) throws SQLException {
  CallableStatement cstmt = conn.prepareCall("CALL sp_create_out_in(?,?)");
  cstmt.registerOutParameter(1, java.sql.Types.INTEGER);
  cstmt.setInt(2, 42);
  cstmt.executeQuery();
  Integer out_value = cstmt.getInt(1);
}
```

## Ejemplos

En el siguiente ejemplo, se llama al nombre del procedimiento `test_spl`.

```
call test_sp1(3,'book');
INFO:  Table "tmp_tbl" does not exist and will be skipped
INFO:  min_val = 3, f2 = book
```

En el siguiente ejemplo, se llama al nombre del procedimiento `test_spl2`.

```
call test_sp2(2,'2019');

         f2          | column2
---------------------+---------
 2019+2019+2019+2019 | 2
(1 row)
```

# CANCEL

Cancela una consulta de base de datos que está actualmente en ejecución.

El comando CANCEL requiere el ID de proceso o el ID de sesión de la consulta en ejecución y muestra un mensaje de confirmación para verificar que se canceló la consulta.

## Privilegios necesarios

Los siguientes privilegios son necesarios para CANCEL:

- Superusuario que cancela su propia consulta
- Superusuario que cancela la consulta de un usuario
- Usuarios con el privilegio CANCEL que cancelan la consulta de un usuario
- Usuario que cancela su propia consulta

## Sintaxis

```
CANCEL process_id [ 'message' ]
```

## Parámetros

process_id

Para cancelar una consulta que se esté ejecutando en un clúster de Amazon Redshift, utilice el `pid` (ID de proceso) de STV_RECENTS que corresponda a la consulta que desee cancelar.

Para cancelar una consulta que se esté ejecutando en un grupo de trabajo de Amazon Redshift sin servidor, utilice el `session_id` de SYS_QUERY_HISTORY que corresponda a la consulta que desee cancelar.

'message'

Un mensaje de confirmación opcional que se muestra cuando se completa la cancelación de la consulta. Si no especifica un mensaje, Amazon Redshift mostrará el mensaje predeterminado como verificación. Debe incluir el mensaje entre comillas simples.

## Notas de uso

No puede cancelar una consulta al especificar un query ID; debe especificar el process ID (PID) o Session ID de la consulta. Solo puede cancelar consultas que su usuario esté ejecutando actualmente. Los superusuarios pueden cancelar todas las consultas.

Si las consultas en varias sesiones mantienen bloqueos en la misma tabla, puede utilizar la función PG_TERMINATE_BACKEND para terminar una de las sesiones. Esta acción obliga a las transacciones en curso de la sesión terminada a eliminar todos los bloqueos y revertir la transacción. Consulte la tabla de sistema STV_LOCKS para ver los bloqueos actuales.

Después de determinados eventos internos, Amazon Redshift podría reiniciar una sesión activa y asignar un PID nuevo. Si el PID ha cambiado, es posible que reciba el siguiente mensaje de error.

```
Session <PID> does not exist. The session PID might have changed. Check the
 stl_restarted_sessions system table for details.
```

Para buscar el nuevo PID, consulte la tabla del sistema STL_RESTARTED_SESSIONS y el filtro en la columna `oldpid`.

```
select oldpid, newpid from stl_restarted_sessions where oldpid = 1234;
```

## Ejemplos

Para cancelar una consulta que se esté ejecutando actualmente en un clúster de Amazon Redshift, primero recupere el ID de proceso para la consulta que desee cancelar. Para determinar los ID de proceso para todas las consultas en ejecución, escriba el siguiente comando:

```
select pid, starttime, duration,
trim(user_name) as user,
trim (query) as querytxt
from stv_recents
where status = 'Running';

pid |         starttime          | duration |   user   |    querytxt
-----+----------------------------+----------+----------+-----------------
802 | 2008-10-14 09:19:03.550885 |      132 | dwuser | select
venuename from venue where venuestate='FL', where venuecity not in
('Miami' , 'Orlando');
834 | 2008-10-14 08:33:49.473585 |  1250414 | dwuser | select *
```

```
from listing;
964 | 2008-10-14 08:30:43.290527 |   326179 | dwuser | select
sellerid from sales where qtysold in (8, 10);
```

Revise el texto de la consulta para determinar cuál es el ID de proceso (PID) que corresponde a la consulta que desea cancelar.

Escriba el siguiente comando para usar PID 802 para cancelar esa consulta:

```
cancel 802;
```

La sesión donde se estaba ejecutando la consulta muestra el siguiente mensaje:

```
ERROR:  Query (168) cancelled on user's request
```

donde 168 es el ID de consulta (no el ID de proceso utilizado para cancelar la consulta).

De manera opcional, puede especificar un mensaje de confirmación personalizado para mostrar en lugar del mensaje predeterminado. Para especificar un mensaje personalizado, incluya el mensaje entre comillas simples al final del comando CANCEL:

```
cancel 802 'Long-running query';
```

La sesión donde se estaba ejecutando la consulta muestra el siguiente mensaje:

```
ERROR:  Long-running query
```

# CLOSE

(Opcional) Cierra todos los recursos libres que están asociados con un cursor abierto. COMMIT, END y ROLLBACK cierran automáticamente el cursor, por lo que no es necesario utilizar el comando CLOSE para cerrarlo explícitamente.

Para obtener más información, consulte DECLARE, FETCH.

## Sintaxis

```
CLOSE cursor
```

### Parámetros

cursor

Nombre del cursor que se cerrará.

### Ejemplo de CLOSE

Los siguientes comandos cierran el cursor y realizan una confirmación, que finaliza la transacción:

```
close movie_cursor;
commit;
```

## COMMENT

Crea o cambia un comentario acerca de un objeto de la base de datos.

### Sintaxis

```
COMMENT ON
{
TABLE object_name |
COLUMN object_name.column_name |
CONSTRAINT constraint_name ON table_name |
DATABASE object_name |
VIEW object_name
}
IS 'text' | NULL
```

### Parámetros

object_name

Nombre del objeto de la base de datos en el que se realiza un comentario. Puede agregar un comentario en los siguientes objetos:

- TABLE
- COLUMN (también lleva un valor column_name [nombre_de_columna]).
- CONSTRAINT (también lleva un valor constraint_name [nombre_de_restricción] y un valor table_name [nombre_de_tabla]).

- DATABASE
- VIEW
- SCHEMA

IS "texto"|NULL

El texto del comentario que desea agregar o sustituir para el objeto especificado. La cadena texto es el tipo de datos TEXT. Incluya el comentario entre comillas simples. Establezca el valor en NULL para eliminar el texto del comentario.

column_name

Nombre de la columna en la que se realiza un comentario. Parámetro de COLUMN. Sigue una tabla especificada en `object_name`.

constraint_name

Nombre de la restricción en la que se realiza un comentario. Parámetro de CONSTRAINT.

table_name

Nombre de una tabla que contiene la restricción. Parámetro de CONSTRAINT.

## Notas de uso

Para agregar o actualizar un comentario, debe ser un superusuario o el propietario de un objeto de base de datos.

Los comentarios en bases de datos solo pueden aplicarse a la base de datos actual. Se muestra un mensaje de advertencia si intenta comentar en una base de datos diferente. Aparece el mismo mensaje de advertencia sobre comentarios de bases de datos que no existen.

No se admiten comentarios en tablas externas, columnas externas ni columnas de vistas de enlace de tiempo de ejecución.

## Ejemplos

En el ejemplo siguiente se agrega un comentario a la tabla SALES.

```
COMMENT ON TABLE sales IS 'This table stores tickets sales data';
```

En el siguiente ejemplo, se muestra el comentario de la tabla SALES.

```
select obj_description('public.sales'::regclass);

obj_description
-----------------------------------
This table stores tickets sales data
```

En el siguiente ejemplo, se elimina un comentario de la tabla SALES.

```
COMMENT ON TABLE sales IS NULL;
```

En el siguiente ejemplo, se agrega un comentario a la columna EVENTID de la tabla SALES.

```
COMMENT ON COLUMN sales.eventid IS 'Foreign-key reference to the EVENT table.';
```

En el siguiente ejemplo, se muestra un comentario en la columna EVENTID (columna número 5) de la tabla SALES.

```
select col_description( 'public.sales'::regclass, 5::integer );

col_description
----------------------------------------
Foreign-key reference to the EVENT table.
```

En el siguiente ejemplo, se agrega un comentario descriptivo a la tabla EVENT.

```
comment on table event is 'Contains listings of individual events.';
```

Consulte el catálogo del sistema PG_DESCRIPTION para ver una lista de comentarios. El siguiente ejemplo devuelve la descripción de la tabla EVENT.

```
select * from pg_catalog.pg_description
where objoid =
(select oid from pg_class where relname = 'event'
and relnamespace =
(select oid from pg_catalog.pg_namespace where nspname = 'public') );
```

```
objoid | classoid | objsubid | description
-------+----------+----------+----------------------------------------
116658 |     1259 |        0 | Contains listings of individual events.
```

# COMMIT

Confirma la transacción actual en la base de datos. Este comando hace que las actualizaciones de transacciones de la base de datos sean permanentes.

## Sintaxis

```
COMMIT [ WORK | TRANSACTION ]
```

## Parámetros

WORK

Palabra clave opcional. Esta palabra clave no se admite en un proceso almacenado.

TRANSACTION

Palabra clave opcional. WORK y TRANSACTION son sinónimos. Ninguna de ellas se admite en un proceso almacenado.

Para más información sobre el uso de COMMIT en un proceso almacenado, vea Administración de transacciones.

## Ejemplos

Cada uno de los siguientes ejemplos confirma la transacción actual en la base de datos:

```
commit;
```

```
commit work;
```

```
commit transaction;
```

# COPY

El cifrado del cliente para los comandos COPY y UNLOAD dejará de estar abierto a nuevos clientes a partir del 30 de abril de 2025. Si ha utilizado el cifrado del cliente con los comandos COPY y UNLOAD en los 12 meses anteriores al 30 de abril de 2025, puede seguir utilizando el cifrado del cliente con los comandos COPY o UNLOAD hasta el 30 de abril de 2026. Después del 30 de abril de 2026, no podrá utilizar el cifrado del cliente para COPY y UNLOAD. Le recomendamos que pase a utilizar el cifrado del servidor para COPY y UNLOAD lo antes posible. Si ya está utilizando el cifrado del servidor para COPY y UNLOAD, no hay ningún cambio y puede seguir utilizándolo sin alterar las consultas. Para obtener más información sobre el cifrado para COPY y UNLOAD, consulte el parámetro ENCRYPTED a continuación.

Carga datos en una tabla desde archivos de datos o desde una tabla de Amazon DynamoDB. Los archivos pueden encontrarse en un bucket de Amazon Simple Storage Service (Amazon S3), un clúster de Amazon EMR o un alojamiento remoto al que se puede acceder mediante una conexión Secure Shell (SSH).

> **Note**
>
> Las tablas externas de Amazon Redshift Spectrum solo se pueden leer. No puede aplicar COPY a una tabla externa.

El comando COPY anexa los datos de entrada como filas adicionales a la tabla.

El tamaño máximo de una fila de entrada única de cualquier origen es de 4 MB.

Temas


- Permisos necesarios
- Sintaxis de COPY
- Parámetros necesarios
- Parámetros opcionales
- Notas de uso y recursos adicionales para el comando COPY
- Ejemplos del comando COPY
- COPY JOB

- Referencia de parámetros de COPY
- Notas de uso
- Ejemplos de COPY

## Permisos necesarios

Para utilizar el comando COPY, debe tener el privilegio INSERT para la tabla de Amazon Redshift.

## Sintaxis de COPY

```
COPY table-name
[ column-list ]
FROM data_source
authorization
[ [ FORMAT ] [ AS ] data_format ]
[ parameter [ argument ] [, ... ] ]
```

Puede realizar una operación COPY con un mínimo de tres parámetros: un nombre de tabla, un origen de datos y una autorización para obtener acceso a los datos.

Amazon Redshift extiende la funcionalidad del comando COPY para permitir la carga de datos en varios formatos de datos de varios orígenes de datos, el control del acceso a los datos de carga, la administración de las transformaciones de datos y la administración de la operación de carga.

Las siguientes secciones presentan los parámetros requeridos del comando COPY, mediante la agrupación de los parámetros opcionales por función. También describen cada parámetro y explican el funcionamiento conjunto de las distintas opciones. Puede ir directamente a la descripción de un parámetro mediante la lista alfabética de parámetros.

## Parámetros necesarios

El comando COPY requiere tres elementos:

- Table Name
- Data Source
- Authorization

El comando COPY más sencillo utiliza el siguiente formato.

```
COPY table-name
FROM data-source
authorization;
```

En el siguiente ejemplo, se crea una tabla denominada CATDEMO y luego se carga la tabla con datos de ejemplo de un archivo de datos de Amazon S3 denominado `category_pipe.txt`.

```
create table catdemo(catid smallint, catgroup varchar(10), catname varchar(10), catdesc
 varchar(50));
```

En el siguiente ejemplo, el origen de datos del comando COPY es un archivo de datos llamado `category_pipe.txt` que se encuentra en la carpeta `tickit` de un bucket de Amazon S3 denominado `redshift-downloads`. El comando COPY tiene autorización para acceder al bucket de Amazon S3 a través de un rol de AWS Identity and Access Management (IAM). Si el clúster tiene adjunto un rol de IAM ya existente con permiso para acceder a Amazon S3, puede sustituir el nombre de recurso de Amazon (ARN) del rol en el siguiente comando COPY y ejecutarlo.

```
copy catdemo
from 's3://redshift-downloads/tickit/category_pipe.txt'
iam_role 'arn:aws:iam::<aws-account-id>:role/<role-name>'
region 'us-east-1';
```

Para obtener instrucciones completas sobre cómo se utilizan los comandos COPY para cargar datos de muestra, incluidas las instrucciones para cargar datos de otras regiones de AWS, consulte Cargar datos de muestra desde Amazon S3 en la Guía de introducción a Amazon Redshift.

table-name (nombre-tabla)

El nombre de la tabla de destino para el comando COPY. La tabla ya debe existir en la base de datos. La tabla puede ser temporal o persistente. El comando COPY adjunta los nuevos datos de entrada a cualquier fila existente en la tabla.

FROM data-source (origen-datos)

La ubicación de los datos de origen que se van a cargar en la tabla de destino. Puede especificarse un archivo de manifiesto con algunos orígenes de datos.

El repositorio de datos más utilizado es un bucket de Amazon S3. También puede cargar datos de archivos de datos ubicados en un clúster de Amazon EMR, una instancia de Amazon EC2 o

un alojamiento remoto al que el clúster puede acceder a través de una conexión SSH, o puede cargarlos directamente desde una tabla de DynamoDB.

- COPY de Amazon S3
- COPY de Amazon EMR
- COPY de hosts remotos (SSH)
- COPY de Amazon DynamoDB

Autorización

Se trata de una cláusula que indica el método que el clúster utiliza para la autenticación y la autorización con objeto de acceder a otros recursos de AWS. El comando COPY necesita autorización para acceder a los datos de otro recurso de AWS, incluidos los recursos de Amazon S3, Amazon EMR, Amazon DynamoDB y Amazon EC2. Puede proporcionar esa autorización haciendo referencia a un rol de IAM que esté asociado al clúster o proporcionando el ID de clave de acceso y la clave de acceso secreta de un usuario de IAM.

- Parámetros de autorización
- Control de acceso con base en roles
- Control de acceso basado en claves

## Parámetros opcionales

De forma opcional, puede especificar el modo en que COPY asigna los datos de campo a las columnas de la tabla de destino, definir los atributos de datos de origen para permitir al comando COPY leer y analizar los datos de origen de forma correcta y administrar qué operaciones realiza el comando COPY durante el proceso de carga.

- Opciones de mapeo de columnas
- Parámetros de formato de datos
- Parámetros de conversión de datos
- Operaciones de carga de datos

Mapeo de columnas

De manera predeterminada, COPY inserta los valores de campo en las columnas de la tabla de destino en el mismo orden que aparecen los campos en los archivos de datos. Si el orden de

columnas predeterminado no funcionará, puede especificar una lista de columnas o utilizar las expresiones JSONPath para asignar los campos de datos de origen a las columnas de destino.

- Column List
- JSONPaths File

### Parámetros de formato de datos

Puede cargar datos de archivos de texto en formato JSON, de valores separados por coma (CSV), de ancho fijo o delimitados por caracteres, o de archivos de Avro.

De manera predeterminada, el comando COPY espera que los datos de origen estén en archivos de texto UTF-8 delimitados por caracteres. El delimitador predeterminado es el carácter de barra vertical ( | ). Si los datos de origen están en otro formato, utilice los siguientes parámetros para especificar el formato de datos.

- FORMAT
- CSV
- DELIMITER
- FIXEDWIDTH
- SHAPEFILE
- AVRO
- JSON
- ENCRYPTED
- BZIP2
- GZIP
- LZOP
- PARQUET
- ORC
- ZSTD

### Parámetros de conversión de datos

A medida que carga la tabla, COPY intenta convertir de forma implícita las cadenas de los datos de origen al tipo de datos de la columna de destino. Si necesita especificar una conversión que sea

diferente a la del comportamiento predeterminado, o si la conversión predeterminada da lugar a errores, puede administrar las conversiones de datos especificando los siguientes parámetros.

- ACCEPTANYDATE
- ACCEPTINVCHARS
- BLANKSASNULL
- DATEFORMAT
- EMPTYASNULL
- ENCODING
- ESCAPE
- EXPLICIT_IDS
- FILLRECORD
- IGNOREBLANKLINES
- IGNOREHEADER
- NULL AS
- REMOVEQUOTES
- ROUNDEC
- TIMEFORMAT
- TRIMBLANKS
- TRUNCATECOLUMNS

Operaciones de carga de datos

Administre el comportamiento predeterminado de la operación de carga para solucionar problemas o reducir los tiempos de carga especificando los siguientes parámetros.

- COMPROWS
- COMPUPDATE
- IGNOREALLERRORS
- MAXERROR
- NOLOAD
- STATUPDATE

## Notas de uso y recursos adicionales para el comando COPY

Para obtener más información acerca de cómo utilizar el comando COPY, consulte los siguientes temas:

- Notas de uso
- Tutorial: Carga de datos desde Amazon S3
- Prácticas recomendadas de Amazon Redshift para la carga de datos
- Carga de tablas con el comando COPY
  - Carga de datos desde Amazon S3
  - Carga de datos desde Amazon EMR
  - Carga de datos desde hosts remotos
  - Carga de datos desde una tabla de Amazon DynamoDB
- Solución de problemas en cargas de datos

## Ejemplos del comando COPY

Para ver más ejemplos que muestran cómo usar COPY desde varios orígenes, en formatos dispares y con diferentes opciones de COPY, consulte Ejemplos de COPY.

## COPY JOB

Para obtener más información acerca del uso de este comando, consulte Creación de una integración de eventos de S3 para copiar automáticamente archivos desde buckets de Amazon S3.

Administra los comandos COPY que cargan datos en una tabla. El comando COPY JOB es una extensión del comando COPY y automatiza la carga de datos desde buckets de Amazon S3. Al crear un trabajo COPY, Amazon Redshift detecta cuándo se crean nuevos archivos de Amazon S3 en una ruta especificada y, a continuación, los carga automáticamente sin su intervención. Al cargar los datos se utilizan los mismos parámetros que en el comando COPY original. Amazon Redshift realiza un seguimiento de los archivos cargados (basándose en el nombre de archivo) para verificar que se cargan solo una vez.

> **Note**
>
> Para obtener información sobre el comando COPY, como su uso, parámetros y permisos, consulte COPY.

Permiso necesario

Para ejecutar el comando COPY de COPY JOB, debe tener privilegios INSERT de la tabla que se va a cargar.

El rol de IAM especificado con el comando COPY debe tener permiso para acceder a los datos que se van a cargar. Para obtener más información, consulte Permisos de IAM para COPY, UNLOAD y CREATE LIBRARY.

Sintaxis

Crear un trabajo de copia. Los parámetros del comando COPY se guardan con el trabajo de copia.

No puede ejecutar COPY JOB CREATE dentro del ámbito de un bloque de transacciones.

```
COPY copy-command JOB CREATE job-name
[AUTO ON | OFF]
```

Cambiar la configuración de un trabajo de copia.

```
COPY JOB ALTER job-name
[AUTO ON | OFF]
```

Ejecutar un trabajo de copia. Se utilizan los parámetros del comando COPY almacenados.

```
COPY JOB RUN job-name
```

Enumerar todos los trabajos de copia.

```
COPY JOB LIST
```

Mostrar los detalles de un trabajo de copia.

```
COPY JOB SHOW job-name
```

Eliminar un trabajo de copia.

No puede ejecutar COPY JOB DROP dentro del ámbito de un bloque de transacciones.

```
COPY JOB DROP job-name
```

Parámetros

copy-command

Un comando COPY que carga datos desde Amazon S3 en Amazon Redshift. La cláusula contiene los parámetros COPY que definen el bucket de Amazon S3, la tabla de destino, el rol de IAM y otros parámetros que se utilizan al cargar datos. Se admiten todos los parámetros del comando COPY para una carga de datos de Amazon S3, con las siguientes excepciones:

- COPY JOB no ingiere archivos preexistentes en la carpeta a la que apunta el comando COPY. Solo se ingieren los archivos creados después de la marca de tiempo de creación de COPY JOB.
- No puede especificar un comando COPY con las opciones MAXERROR o IGNOREALLERRORS.
- No puede especificar un archivo de manifiesto. COPY JOB requiere una ubicación de Amazon S3 designada para supervisar los archivos recién creados.
- No puede especificar un comando COPY con tipos de autorización como Acceso y Claves secretas. Solo se admiten los comandos COPY que utilizan el parámetro `IAM_ROLE` para la autorización. Para obtener más información, consulte Parámetros de autorización.
- COPY JOB no admite el rol de IAM predeterminado asociado al clúster. Debe especificar `IAM_ROLE` en el comando COPY.

Para obtener más información, consulte COPY de Amazon S3.

job-name

El nombre del trabajo que se usa para hacer referencia al trabajo COPY. El valor de job-name no puede contener un guion (-).

[AUTO ON | OFF]

Cláusula que indica si los datos de Amazon S3 se cargan automáticamente en las tablas de Amazon Redshift.

- Si es `ON`, Amazon Redshift supervisa la ruta de origen de Amazon S3 en busca de archivos recién creados y, si los encuentra, se ejecuta un comando COPY con los parámetros de COPY en la definición del trabajo. Esta es la opción predeterminada.
- Si es `OFF`, Amazon Redshift no ejecuta COPY JOB automáticamente.

Notas de uso

Las opciones del comando COPY no se validan hasta el momento de la ejecución. Por ejemplo, un `IAM_ROLE` no válido o un origen de datos de Amazon S3 provoca errores de ejecución cuando se inicia COPY JOB.

Si el clúster está en pausa, no se ejecutan los comandos COPY JOB.

Para consultar los archivos del comando COPY cargados y los errores de carga, consulte STL_LOAD_COMMITS, STL_LOAD_ERRORS y STL_LOADERROR_DETAIL. Para obtener más información, consulte Comprobación de carga correcta de datos.

Ejemplos

En el siguiente ejemplo se muestra la creación de un trabajo COPY JOB para cargar datos de un bucket de Amazon S3.

```
COPY public.target_table
FROM 's3://amzn-s3-demo-bucket/staging-folder'
IAM_ROLE 'arn:aws:iam::123456789012:role/MyLoadRoleName'
JOB CREATE my_copy_job_name
AUTO ON;
```

## Referencia de parámetros de COPY

COPY tiene muchos parámetros que se pueden utilizar en muchas situaciones. Sin embargo, no todos los parámetros son compatibles en cada situación. Por ejemplo, para cargar desde archivos ORC o PARQUET, se admite un número limitado de parámetros. Para obtener más información, consulte Uso de COPY con formatos de datos de columnas.

Temas

- Origen de datos
- Parámetros de autorización
- Opciones de mapeo de columnas
- Parámetros de formato de datos
- Parámetros de compresión de archivos
- Parámetros de conversión de datos
- Operaciones de carga de datos

- Lista alfabética de parámetros

### Origen de datos

Puede cargar datos de archivos de texto en un bucket de Amazon S3, un clúster de Amazon EMR o un alojamiento remoto al cual el clúster pueda acceder a través de una conexión SSH. También puede cargar datos de forma directa desde una tabla de DynamoDB.

El tamaño máximo de una fila de entrada única de cualquier origen es de 4 MB.

Para exportar los datos de una tabla a un conjunto de archivos en Amazon S3, utilice el comando UNLOAD.

Temas

- COPY de Amazon S3
- COPY de Amazon EMR
- COPY de hosts remotos (SSH)
- COPY de Amazon DynamoDB

### COPY de Amazon S3

Para cargar datos de archivos ubicados en uno o más buckets de S3, utilice la cláusula FROM para indicar el modo en que COPY localiza los archivos de Amazon S3. Puede proporcionar la ruta de objeto a los archivos de datos como parte de la cláusula FROM o puede proporcionar la ubicación de un archivo de manifiesto que contenga una lista de rutas de objetos de Amazon S3. COPY de Amazon S3 utiliza una conexión HTTPS. Asegúrese de que los rangos de IP de S3 estén agregados a la lista de permitidos. Para obtener más información acerca de los rangos de IP de S3 necesarios, consulte Aislamiento de red.

> **Important**
>
> Si los buckets de Amazon S3 que contienen los archivos de datos no se encuentran en la misma región de AWS que el clúster, debe utilizar el parámetro REGION para especificar la región en la que se encuentran los datos.

Temas

- Sintaxis
- Ejemplos
- Parámetros opcionales
- Parámetros no admitidos


### Sintaxis

```
FROM { 's3://objectpath' | 's3://manifest_file' }
authorization
| MANIFEST
| ENCRYPTED
| REGION [AS] 'aws-region'
| optional-parameters
```

### Ejemplos

En el siguiente ejemplo, se utiliza una ruta de objeto para cargar datos desde Amazon S3.

```
copy customer
from 's3://amzn-s3-demo-bucket/customer'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole';
```

En el siguiente ejemplo, se utiliza un archivo de manifiesto para cargar datos desde Amazon S3.

```
copy customer
from 's3://amzn-s3-demo-bucket/cust.manifest'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
manifest;
```

### Parámetros

FROM

El origen de los datos a cargar. Para obtener más información acerca de la codificación del archivo de Amazon S3, consulte Parámetros de conversión de datos.

's3://copy_from_s3_objectpath'

Especifica la ruta a los objetos de Amazon S3 que contienen los datos; por ejemplo, `'s3://amzn-s3-demo-bucket/custdata.txt'`. El parámetro s3://copy_from_s3_objectpath

puede hacer referencia a un único archivo o a un conjunto de objetos o carpetas que tienen el mismo prefijo de clave. Por ejemplo, el nombre `custdata.txt` es un prefijo de clave que hace referencia a un número de archivos físicos: `custdata.txt`, `custdata.txt.1`, `custdata.txt.2`, `custdata.txt.bak` y así sucesivamente. El prefijo de clave también puede hacer referencia a un número de carpetas. Por ejemplo, `'s3://amzn-s3-demo-bucket/custfolder'` hace referencia a las carpetas `custfolder`, `custfolder_1`, `custfolder_2` y así sucesivamente. Si un prefijo de clave hace referencia a varias carpetas, se cargan todos los archivos de esas carpetas. Si un prefijo de clave coincide con un archivo y, a su vez, con una carpeta, como `custfolder.log`, COPY también intenta cargar el archivo. Si un prefijo de clave puede dar lugar a que COPY intente cargar archivos no deseados, utilice un archivo de manifiesto. Para obtener más información, consulte copy_from_s3_manifest_file, a continuación.

> **Important**
>
> Si el bucket de S3 que contiene los archivos de datos no se encuentra en la misma región de AWS que el clúster, debe utilizar el parámetro REGION para especificar la región en la que se encuentran los datos.

Para obtener más información, consulte Carga de datos desde Amazon S3.

's3://copy_from_s3_manifest_file'

Especifica la clave de objeto de Amazon S3 para un archivo de manifiesto que muestra los archivos de datos que se cargarán. El argumento 's3://copy_from_s3_manifest_file' debe referenciar de forma explícita un solo archivo; por ejemplo, `'s3://amzn-s3-demo-bucket/manifest.txt'`. No puede hacer referencia a un prefijo de clave.

El manifiesto es un archivo de texto en formato JSON que muestra la dirección URL de cada archivo que se cargará desde Amazon S3. El URL incluye el nombre del bucket y la ruta de objeto completa para el archivo. Los archivos que se especifican en el manifiesto pueden estar en buckets diferentes, pero todos los buckets deben estar en la misma región de AWS que el clúster de Amazon Redshift. Si un archivo aparece dos veces, este se carga dos veces. En el siguiente ejemplo, se muestra el JSON para un manifiesto que carga tres archivos.

```
{
  "entries": [
    {"url":"s3://amzn-s3-demo-bucket1/custdata.1","mandatory":true},
    {"url":"s3://amzn-s3-demo-bucket1/custdata.2","mandatory":true},
```

```
    {"url":"s3://amzn-s3-demo-bucket2/custdata.1","mandatory":false}
  ]
}
```

Se requieren caracteres de comillas dobles y estas deben ser las comillas simples (0x22), no las comillas inclinadas o “inteligentes”. Cada entrada en el manifiesto puede incluir opcionalmente una marca `mandatory`. Cuando `mandatory` está establecido en `true`, COPY termina si no encuentra el archivo de esa entrada. De lo contrario, COPY continúa. El valor predeterminado de `mandatory` es `false`.

Cuando la carga se realiza a partir de archivos de datos con formato ORC o Parquet, se necesita un campo `meta`, tal y como se muestra en el siguiente ejemplo.

```
{
   "entries":[
      {
         "url":"s3://amzn-s3-demo-bucket1/orc/2013-10-04-custdata",
         "mandatory":true,
         "meta":{
            "content_length":99
         }
      },
      {
         "url":"s3://amzn-s3-demo-bucket2/orc/2013-10-05-custdata",
         "mandatory":true,
         "meta":{
            "content_length":99
         }
      }
   ]
}
```

El archivo de manifiesto no debe estar cifrado o comprimido, incluso si se especifican las opciones ENCRYPTED, GZIP, LZOP, BZIP2 o ZSTD. COPY devuelve un error si el archivo de manifiesto especificado no se encuentra o no se creó de forma correcta.

Si se utiliza un archivo de manifiesto, se debe especificar el parámetro MANIFEST con el comando COPY. Si no se especifica el parámetro MANIFEST, COPY supone que el archivo especificado con FROM es un archivo de datos.

Para obtener más información, consulte Carga de datos desde Amazon S3.

authorization

El comando COPY necesita autorización para acceder a los datos de otro recurso de AWS, incluidos los recursos de Amazon S3, Amazon EMR, Amazon DynamoDB y Amazon EC2. Puede proporcionar esa autorización referenciando un rol de AWS Identity and Access Management (IAM) que esté adjunto al clúster (control de acceso basado en roles) o proporcionando las credenciales de acceso de un usuario (control de acceso basado en claves). Para mayor seguridad y flexibilidad, le recomendamos utilizar un control de acceso basado en roles de IAM. Para obtener más información, consulte Parámetros de autorización.

MANIFEST

Especifica que se utiliza un manifiesto para identificar los archivos de datos que se cargarán desde Amazon S3. Si se utiliza el parámetro MANIFEST, COPY carga los datos de los archivos que se muestran en el manifiesto al que se hace referencia en 's3://copy_from_s3_manifest_file'. Si no se encuentra el archivo de manifiesto o no se creó de forma correcta, COPY no se ejecuta correctamente. Para obtener más información, consulte Uso de un manifiesto para especificar archivos de datos.

ENCRYPTED

Se trata de una cláusula que especifica que los archivos de entrada en Amazon S3 están cifrados con el cifrado del lado del cliente con claves administradas por el cliente. Para obtener más información, consulte Carga de archivos de datos cifrados desde Amazon S3. No especifique ENCRYPTED si los archivos de entrada están cifrados con el cifrado de servidor de Amazon S3 (SSE-KMS o SSE-S3). COPY lee automáticamente los archivos con cifrados del servidor.

Si especifica el parámetro ENCRYPTED, también debe especificar el parámetro MASTER_SYMMETRIC_KEY o incluir el valor **`master_symmetric_key`** en la cadena the section called “Uso de CREDENTIALS”.

Si los archivos cifrados están en un formato comprimido, agregue el parámetro GZIP, LZOP, BZIP2 o ZSTD.

Los archivos de manifiesto y los archivos JSONPaths no deben cifrarse aunque la opción ENCRYPTED esté especificada.

MASTER_SYMMETRIC_KEY 'root_key'

Se trata de la clave raíz simétrica que se utilizó para cifrar los archivos de datos en Amazon S3. Si se especifica MASTER_SYMMETRIC_KEY, también se debe especificar el

parámetro ENCRYPTED. MASTER_SYMMETRIC_KEY no puede utilizarse con el parámetro CREDENTIALS. Para obtener más información, consulte Carga de archivos de datos cifrados desde Amazon S3.

Si los archivos cifrados están en un formato comprimido, agregue el parámetro GZIP, LZOP, BZIP2 o ZSTD.

REGION [AS] 'aws_region'

Especifica la región de AWS en la que se encuentran los datos de origen. Cuando el recurso de AWS que contiene los datos no se encuentra en la misma región que el clúster de Amazon Redshift, se debe utilizar REGION para el comando COPY de un bucket de Amazon S3 o una tabla de DynamoDB

El valor de aws_region debe coincidir con una de las regiones que aparecen en la tabla Regiones y puntos de enlace de Amazon Redshift.

Si se especifica el parámetro REGION, todos los recursos, incluidos un archivo de manifiesto o varios buckets de Amazon S3, deben encontrarse en la región especificada.

> **Note**
>
> La transferencia de datos entre regiones genera cargos adicionales en el bucket de Amazon S3 o la tabla de DynamoDB que contiene los datos. Para obtener más información acerca de los precios, consulte Transferencia SALIENTE de datos de Amazon S3 a otra región de AWS en la página Precios de Amazon S3 y Transferencia SALIENTE de datos en la página Precios de Amazon DynamoDB.

De manera predeterminada, COPY supone que los datos se encuentran en la misma región que el clúster de Amazon Redshift.

Parámetros opcionales

Si lo prefiere, puede especificar los siguientes parámetros con COPY de Amazon S3:

- Opciones de mapeo de columnas
- Parámetros de formato de datos
- Parámetros de conversión de datos

- Operaciones de carga de datos

### Parámetros no admitidos

No puede utilizar los siguientes parámetros con COPY de Amazon S3:

- SSH
- READRATIO

## COPY de Amazon EMR

Puede utilizar el comando COPY para cargar datos en paralelo desde un clúster de Amazon EMR configurado para escribir archivos de texto en el sistema de archivos Hadoop Distributed File System (HDFS) del clúster como archivos de ancho fijo, archivos delimitados por caracteres, archivos CSV, archivos con formato JSON o archivos Avro.

Temas

- Sintaxis
- Ejemplo
- Parámetros
- Parámetros admitidos
- Parámetros no admitidos

### Sintaxis

```
FROM 'emr://emr_cluster_id/hdfs_filepath'
authorization
[ optional_parameters ]
```

### Ejemplo

En el siguiente ejemplo, se cargan datos desde un clúster de Amazon EMR.

```
copy sales
from 'emr://j-SAMPLE2B500FC/myoutput/part-*'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole';
```

Parámetros

FROM

El origen de los datos a cargar.

'emr://emr_cluster_id/hdfs_file_path'

Se trata del identificador único para el clúster de Amazon EMR y la ruta del archivo de HDFS que referencia los archivos de datos para el comando COPY. Los nombres de los archivos de datos HDFS no deben contener los caracteres comodín de asterisco (*) ni de signos de interrogación (?).

> **Note**
>
> El clúster de Amazon EMR debe seguir en ejecución hasta que se complete la operación COPY. Si se cambia o se elimina cualquiera de los archivos de datos HDFS antes de que se complete la operación COPY, se pueden obtener resultados inesperados o la operación COPY puede fallar.

Puede utilizar los caracteres comodín asterisco (*) y signo de puntuación (?) en el argumento hdfs_file_path para especificar que se carguen varios archivos. Por ejemplo, `'emr://j-SAMPLE2B500FC/myoutput/part*'` identifica los archivos `part-0000`, `part-0001` y así sucesivamente. Si la ruta de archivo no contiene caracteres comodín, se trata como si fuera un literal de cadena. Si solo especifica el nombre de una carpeta, COPY prueba cargar todos los archivos que se encuentran en ella.

> **Important**
>
> Si utiliza caracteres comodín o solo el nombre de la carpeta, verifique que no se cargarán archivos no deseados. Por ejemplo, algunos procesos podrían escribir un archivo de registro en la carpeta de salida.

Para obtener más información, consulte Carga de datos desde Amazon EMR.

authorization

El comando COPY necesita autorización para acceder a los datos de otro recurso de AWS, incluidos los recursos de Amazon S3, Amazon EMR, Amazon DynamoDB y Amazon EC2. Puede

proporcionar esa autorización referenciando un rol de AWS Identity and Access Management (IAM) que esté adjunto al clúster (control de acceso basado en roles) o proporcionando las credenciales de acceso de un usuario (control de acceso basado en claves). Para mayor seguridad y flexibilidad, le recomendamos utilizar un control de acceso basado en roles de IAM. Para obtener más información, consulte Parámetros de autorización.

Parámetros admitidos

Si lo prefiere, puede especificar los siguientes parámetros con COPY de Amazon EMR:

- Opciones de mapeo de columnas
- Parámetros de formato de datos
- Parámetros de conversión de datos
- Operaciones de carga de datos

Parámetros no admitidos

No puede utilizar los siguientes parámetros con COPY de Amazon EMR:

- ENCRYPTED
- MANIFEST
- REGION
- READRATIO
- SSH

COPY de hosts remotos (SSH)

Puede utilizar el comando COPY para cargar datos en paralelo desde uno o más alojamientos remotos, como instancias de Amazon Elastic Compute Cloud (Amazon EC2) u otros equipos. COPY se conecta a los hosts remotos mediante Secure Shell (SSH) y ejecuta los comandos en los hosts remotos para generar la salida de texto. El alojamiento remoto puede ser una instancia EC2 de Linux u otro equipo Linux o Unix configurada para aceptar conexiones SSH. Amazon Redshift puede conectarse a varios alojamientos y puede establecer varias conexiones SSH en cada alojamiento. Amazon Redshift envía un comando único a través de cada conexión para generar una salida de texto que aparecerá en la salida estándar del alojamiento, la cual Amazon Redshift leerá después como si fuese cualquier otro archivo de texto.

Utilice la cláusula FROM para especificar la clave de objeto de Amazon S3 para el archivo de manifiesto que proporciona la información que COPY utilizará para establecer las conexiones SSH y ejecutar los comandos remotos.

Temas


- Sintaxis
- Ejemplos
- Parámetros
- Parámetros opcionales
- Parámetros no admitidos


> **Important**
>
> Si el bucket de S3 que contiene el archivo de manifiesto no se encuentra en la misma región de AWS que el clúster, debe utilizar el parámetro REGION para especificar la región en la que se encuentra el bucket.

## Sintaxis

```
FROM 's3://'ssh_manifest_file' }
authorization
SSH
| optional-parameters
```

## Ejemplos

En el siguiente ejemplo, se utiliza un archivo de manifiesto para cargar datos desde un host remoto mediante SSH.

```
copy sales
from 's3://amzn-s3-demo-bucket/ssh_manifest'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
ssh;
```

Parámetros

FROM

El origen de los datos a cargar.

's3://copy_from_ssh_manifest_file'

El comando COPY puede conectarse con distintos hosts mediante Secure Shell (SSH, Shell seguro) y puede crear distintas conexiones SSH a cada host. COPY ejecuta un comando a través de cada conexión al host y, luego, carga la salida en la tabla desde los comandos en paralelo. El argumento s3://copy_from_ssh_manifest_file especifica la clave de objeto de Amazon S3 del archivo de manifiesto que proporciona la información que COPY utiliza para establecer conexiones SSH y ejecutar los comandos remotos.

El argumento s3://copy_from_ssh_manifest_file debe referenciar de forma explícita un solo archivo y no puede ser un prefijo de clave. A continuación se muestra un ejemplo:

```
's3://amzn-s3-demo-bucket/ssh_manifest.txt'
```

El archivo de manifiesto es un archivo de texto con formato JSON que Amazon Redshift usa para conectarse al alojamiento. En el archivo de manifiesto, se especifican los puntos de conexión del host SSH y los comandos que se ejecutarán en los hosts para devolver datos a Amazon Redshift. De forma opcional, puede incluir la clave pública del host, el nombre de usuario de inicio de sesión y una marca obligatoria para cada entrada. En el siguiente ejemplo, se muestra un archivo de manifiesto que crea dos conexiones SSH:

```
{
    "entries": [
     {"endpoint":"<ssh_endpoint_or_IP>",
           "command": "<remote_command>",
           "mandatory":true,
           "publickey": "<public_key>",
           "username": "<host_user_name>"},
     {"endpoint":"<ssh_endpoint_or_IP>",
           "command": "<remote_command>",
           "mandatory":true,
           "publickey": "<public_key>",
           "username": "<host_user_name>"}
     ]
}
```

El archivo de manifiesto contiene una construcción `"entries"` para cada conexión SSH. Puede tener distintas conexiones a un único host o distintas conexiones a distintos hosts. Se requieren caracteres de comillas dobles, como se muestra, para los valores y los nombres de campo. Las comillas deben ser las comillas simples (0x22), no las comillas inclinadas o “inteligentes”. El único valor que no necesita caracteres de comillas dobles es el valor booleano `true` o `false` del campo `"mandatory"`.

En la siguiente lista se describen los campos del archivo de manifiesto.

punto de conexión

Se trata de la dirección URL o la dirección IP del alojamiento; por ejemplo, `"ec2-111-222-333.compute-1.amazonaws.com"` o `"198.51.100.0"`.

comando

El comando que ejecutará el host para generar la salida de texto o la salida binaria en formato gzip, lzop, bzip2 o zstd. El comando puede ser cualquiera que el usuario "host_user_name" (nombre_de_usuario_del_host) tenga permiso para ejecutar. El comando puede ser tan sencillo como imprimir un archivo, o puede consultar una base de datos o lanzar un script. La salida (archivo de texto o archivos binarios gzip, lzop o bzip2) debe estar en un formato que el comando COPY de Amazon Redshift pueda capturar. Para obtener más información, consulte Preparación de los datos de entrada.

publickey

(Opcional) La clave pública del host. Si se proporciona la clave pública, Amazon Redshift la usará para identificar el alojamiento. Si no se proporciona la clave pública, Amazon Redshift no intentará identificar el alojamiento. Por ejemplo, si la clave pública del host remoto es `ssh-rsa AbcCbaxxx…Example root@amazon.com`, escriba el siguiente texto en el campo de clave pública: `"AbcCbaxxx…Example"`.

mandatory

(Opcional) Una cláusula que indica si el comando COPY debe fallar en caso de que el intento de conexión falle. El valor predeterminado es `false`. Si Amazon Redshift no logra establecer al menos una conexión, el comando COPY presentará error.

nombre de usuario

(Opcional) Se trata del nombre de usuario que se utilizará para iniciar sesión en el sistema de alojamiento y ejecutar el comando remoto. El nombre de inicio de sesión del usuario debe ser

el mismo que se utilizó para agregar la clave pública del clúster de Amazon Redshift al archivo de claves autorizadas del alojamiento. El nombre de usuario predeterminado es `redshift`.

Para obtener más información acerca de la creación de un archivo de manifiesto, consulte Proceso de carga de datos.

Para utilizar COPY desde un host remoto, se debe especificar el parámetro SSH con el comando COPY. Si no se especifica el parámetro SSH, COPY supone que el archivo especificado con FROM es un archivo de datos y no se ejecutará correctamente.

Si utiliza la compresión automática, el comando COPY realiza dos operaciones de lectura de los datos, lo que significa que ejecutará el comando remoto dos veces. La primera operación de lectura se realiza para proporcionar una muestra de datos para el análisis de compresión, la segunda operación de lectura es la que carga los datos. Si ejecutar el comando remoto dos veces puede producir un problema, debe deshabilitar la compresión automática. Para deshabilitar la compresión automática, ejecute el comando COPY con el parámetro COMPUPDATE establecido en OFF. Para obtener más información, consulte Carga de tablas con compresión automática.

Para obtener los procedimientos detallados para utilizar COPY de SSH, consulte Carga de datos desde hosts remotos.

authorization

El comando COPY necesita autorización para acceder a los datos de otro recurso de AWS, incluidos los recursos de Amazon S3, Amazon EMR, Amazon DynamoDB y Amazon EC2. Puede proporcionar esa autorización referenciando un rol de AWS Identity and Access Management (IAM) que esté adjunto al clúster (control de acceso basado en roles) o proporcionando las credenciales de acceso de un usuario (control de acceso basado en claves). Para mayor seguridad y flexibilidad, le recomendamos utilizar un control de acceso basado en roles de IAM. Para obtener más información, consulte Parámetros de autorización.

SSH

Una cláusula que especifica que los datos se van a cargar desde un host remoto mediante el protocolo SSH. Si especifica SSH, también debe proporcionar un archivo de manifiesto mediante el argumento s3://copy_from_ssh_manifest_file.

> **Note**
>
> Si utiliza SSH para copiar datos de un host con una dirección IP privada en una VPC remota, la VPC debe tener habilitado el direccionamiento de VPC mejorado. Para obtener

más información sobre el direccionamiento de VPC mejorado, consulte el artículo sobre el direccionamiento de VPC mejorado en Amazon Redshift

### Parámetros opcionales

Opcionalmente, puede especificar los siguientes parámetros con COPY de SSH:

- Opciones de mapeo de columnas
- Parámetros de formato de datos
- Parámetros de conversión de datos
- Operaciones de carga de datos

### Parámetros no admitidos

No puede utilizar los siguientes parámetros con COPY de SSH:

- ENCRYPTED
- MANIFEST
- READRATIO

## COPY de Amazon DynamoDB

Si desea cargar datos de una tabla existente de DynamoDB, utilice la cláusula FROM para especificar el nombre de la tabla de DynamoDB.

Temas


- Sintaxis
- Ejemplos
- Parámetros opcionales
- Parámetros no admitidos

> **Important**
>
> Si la tabla de DynamoDB no se encuentra en la misma región que el clúster de Amazon Redshift, debe utilizar el parámetro REGION para especificar la región en la que se encuentran los datos.

### Sintaxis

```
FROM 'dynamodb://table-name'
authorization
READRATIO ratio
| REGION [AS] 'aws_region'
| optional-parameters
```

### Ejemplos

En el siguiente ejemplo, se cargan datos de una tabla de DynamoDB.

```
copy favoritemovies from 'dynamodb://ProductCatalog'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
readratio 50;
```

### Parámetros

FROM

El origen de los datos a cargar.

'dynamodb://table-name (nombre-tabla)'

Se trata del nombre de la tabla de DynamoDB que contiene los datos; por ejemplo, `'dynamodb://ProductCatalog'`. Para obtener detalles acerca de cómo se asignan los atributos de DynamoDB a las columnas de Amazon Redshift, consulte Carga de datos desde una tabla de Amazon DynamoDB.

El nombre de una tabla de DynamoDB es único para una cuenta de AWS, que se identifica con las credenciales de acceso de AWS.

authorization

El comando COPY necesita autorización para acceder a los datos de otro recurso de AWS, incluidos los recursos de Amazon S3, Amazon EMR, DynamoDB y Amazon EC2. Puede proporcionar esa autorización referenciando un rol de AWS Identity and Access Management (IAM) que esté adjunto al clúster (control de acceso basado en roles) o proporcionando las credenciales de acceso de un usuario (control de acceso basado en claves). Para mayor seguridad y flexibilidad, le recomendamos utilizar un control de acceso basado en roles de IAM. Para obtener más información, consulte Parámetros de autorización.

READRATIO [AS] ratio

Se trata del porcentaje del rendimiento aprovisionado de la tabla de DynamoDB para utilizar en la carga de datos. Se requiere READRATIO para utilizar COPY de DynamoDB. No se puede utilizar con COPY de Amazon S3. Le recomendamos encarecidamente establecer la ratio en un valor menor que el rendimiento provisionado sin utilizar promedio. Los valores válidos son números enteros que forman parte del rango 1-200.

> **Important**
>
> Si establece READRATIO en 100 o más permite a Amazon Redshift consumir la totalidad del rendimiento aprovisionado de la tabla de DynamoDB, lo que degrada seriamente el rendimiento de las operaciones de lectura simultáneas que se produzcan en la misma tabla durante la sesión de COPY. El tráfico de escritura no resulta afectado. Los valores superiores a 100 se permiten para solucionar problemas en situaciones raras por las que Amazon Redshift no cumple el rendimiento aprovisionado de la tabla. Si carga datos de DynamoDB en Amazon Redshift de forma continua, considere la posibilidad de organizar las tablas de DynamoDB como series temporales para separar el tráfico activo de la operación COPY.

Parámetros opcionales

Si lo prefiere, puede especificar los siguientes parámetros con COPY de Amazon DynamoDB:

- Opciones de mapeo de columnas
- Se admiten los siguientes parámetros de conversión de datos:
  - ACCEPTANYDATE
  - BLANKSASNULL

  - DATEFORMAT
  - EMPTYASNULL
  - ROUNDEC
  - TIMEFORMAT
  - TRIMBLANKS
  - TRUNCATECOLUMNS
- Operaciones de carga de datos

Parámetros no admitidos

No se pueden utilizar los siguientes parámetros con COPY de DynamoDB:

- Todos los parámetros de formatos de datos
- ESCAPE
- FILLRECORD
- IGNOREBLANKLINES
- IGNOREHEADER
- NULL
- REMOVEQUOTES
- ACCEPTINVCHARS
- MANIFEST
- ENCRYPTED

Parámetros de autorización

El comando COPY necesita autorización para acceder a los datos de otro recurso de AWS, incluidos los recursos de Amazon S3, Amazon EMR, Amazon DynamoDB y Amazon EC2. Puede proporcionar esa autorización si hace referencia a un rol de (IAM) de AWS Identity and Access Management que esté asociado a su clúster (role-based access control). Puede cifrar los datos de carga en Amazon S3.

En los siguientes temas se proporcionan más detalles y ejemplos de opciones de autenticación:

- Permisos de IAM para COPY, UNLOAD y CREATE LIBRARY

- Control de acceso con base en roles
- Control de acceso basado en claves

Utilice una de las siguientes opciones para brindar autorización al comando COPY:

- Parámetro the section called “Uso de IAM_ROLE”
- Parámetros the section called “Uso de ACCESS_KEY_ID y SECRET_ACCESS_KEY”
- Cláusula the section called “Uso de CREDENTIALS”

Uso del parámetro IAM_ROLE

IAM_ROLE

Utilice la palabra clave predeterminada para que Amazon Redshift utilice el rol de IAM configurado como predeterminado y asociado al clúster cuando se ejecuta el comando COPY.

Utilice el nombre de recurso de Amazon (ARN), de un rol de IAM que el clúster utiliza para la autenticación y la autorización. Si especifica IAM_ROLE, no puede utilizar ACCESS_KEY_ID, SECRET_ACCESS_KEY, SESSION_TOKEN ni CREDENTIALS.

A continuación, se muestra la sintaxis del parámetro IAM_ROLE.

```
IAM_ROLE { default | 'arn:aws:iam::<Cuenta de AWS-id>:role/<role-name>' }
```

Para obtener más información, consulte Control de acceso con base en roles.

Uso de los parámetros ACCESS_KEY_ID y SECRET_ACCESS_KEY

ACCESS_KEY_ID, SECRET_ACCESS_KEY

No se recomienda este método de autorización.

> **Note**
>
> En lugar de proporcionar credenciales de acceso como texto sin formato, le recomendamos encarecidamente utilizar una autenticación basada en roles mediante la especificación del parámetro IAM_ROLE. Para obtener más información, consulte Control de acceso con base en roles.

SESSION_TOKEN

El token de sesión para utilizar con credenciales de acceso temporales. Cuando se especifica SESSION_TOKEN, también debe utilizar ACCESS_KEY_ID y SECRET_ACCESS_KEY para proporcionar credenciales temporales de clave de acceso. Si especifica SESSION_TOKEN, no puede utilizar IAM_ROLE ni CREDENTIALS. Para obtener más información, consulte Credenciales de seguridad temporales en la Guía del usuario de IAM.

> **Note**
>
> En lugar de crear credenciales de seguridad temporales, le recomendamos encarecidamente utilizar una autenticación basada en roles. Cuando brinda autorización mediante un rol de IAM, Amazon Redshift crea de forma automática credenciales temporales de usuario para cada sesión. Para obtener más información, consulte Control de acceso con base en roles.

A continuación se muestra la sintaxis del parámetro SESSION_TOKEN con los parámetros ACCESS_KEY_ID y SECRET_ACCESS_KEY.

```
ACCESS_KEY_ID '<access-key-id>'
SECRET_ACCESS_KEY '<secret-access-key>'
SESSION_TOKEN '<temporary-token>';
```

Si especifica SESSION_TOKEN, no puede utilizar CREDENTIALS ni IAM_ROLE.

Uso del parámetro CREDENTIALS

CREDENTIALS

Se trata de una cláusula que indica el método que el clúster utilizará cuando obtenga acceso a otros recursos de AWS que contienen archivos de datos o archivos de manifiesto. No puede utilizar el parámetro CREDENTIALS con IAM_ROLE ni con ACCESS_KEY_ID y SECRET_ACCESS_KEY.

A continuación, se muestra la sintaxis del parámetro CREDENTIALS.

```
[WITH] CREDENTIALS [AS] 'credentials-args'
```

> **Note**
>
> Para mayor flexibilidad, recomendamos utilizar el parámetro IAM_ROLE en lugar del parámetro CREDENTIALS.

De manera opcional, si se usa el parámetro ENCRYPTED, la cadena credentials-args (credenciales-argumentos) también proporciona la clave de cifrado.

La cadena credentials-args distingue entre mayúsculas y minúsculas y no debe contener espacios.

Las palabras clave WITH y AS son opcionales y se ignoran.

Puede especificar role-based access control o key-based access control. En ambos casos, el rol o el usuario de IAM deben tener los permisos necesarios para acceder a los recursos de AWS especificados. Para obtener más información, consulte Permisos de IAM para COPY, UNLOAD y CREATE LIBRARY.

> **Note**
>
> Para salvaguardar las credenciales de AWS y proteger la información confidencial, le recomendamos encarecidamente utilizar un control de acceso basado en roles.

Para especificar un control de acceso basado en funciones, proporcione la cadena credentials-args (credenciales-argumentos) en el siguiente formato.

```
'aws_iam_role=arn:aws:iam::<aws-account-id>:role/<role-name>'
```

Para usar las credenciales de token temporales, debe proporcionar el ID de clave de acceso temporal, la clave de acceso secreta temporal y el token temporal. La cadena credentials-args tiene el siguiente formato.

```
CREDENTIALS
'aws_access_key_id=<temporary-access-key-id>;aws_secret_access_key=<temporary-secret-access-key>;token=<temporary-token>'
```

Un comando COPY que utilice un control de acceso basado en roles con credenciales temporales se parecería a la siguiente instrucción de ejemplo:

```
COPY customer FROM 's3://amzn-s3-demo-bucket/mydata'
CREDENTIALS
'aws_access_key_id=<temporary-access-key-id>;aws_secret_access_key=<temporary-secret-access-key-id>;token=<temporary-token>'
```

Para obtener más información, consulte Credenciales de seguridad temporales.

Si se utiliza el parámetro ENCRYPTED, la cadena credentials-args tiene el siguiente formato, donde *<root-key>* es el valor de la clave raíz que se utilizó para cifrar los archivos.

```
CREDENTIALS
'<credentials-args>;master_symmetric_key=<root-key>'
```

Un comando COPY que utilice un control de acceso basado en roles con una clave de cifrado se parecería a la siguiente instrucción de ejemplo:

```
COPY customer FROM 's3://amzn-s3-demo-bucket/mydata'
CREDENTIALS
'aws_iam_role=arn:aws:iam::<account-id>:role/<role-name>;master_symmetric_key=<root-key>'
```

Opciones de mapeo de columnas

De manera predeterminada, COPY inserta los valores en las columnas de la tabla de destino en el mismo orden que aparecen los campos en los archivos de datos. Si el orden de columnas predeterminado no funcionará, puede especificar una lista de columnas o utilizar las expresiones JSONPath para asignar los campos de datos de origen a las columnas de destino.

- Column List
- JSONPaths File

Lista de columnas

Puede especificar una lista de nombres de columnas separada por comas para cargar los campos de datos de origen en columnas de destino específicas. Las columnas pueden estar en cualquier orden en la instrucción COPY, pero, cuando se cargan desde archivos sin formato, como en un bucket de Amazon S3, el orden debe coincidir con el de los datos de origen.

Cuando se realiza la carga desde una tabla de Amazon DynamoDB, el orden no importa. El comando COPY establece una correspondencia entre los nombres de atributos de los elementos recuperados

de la tabla de DynamoDB y los nombres de columnas de una tabla de Amazon Redshift. Para obtener más información, consulte Carga de datos desde una tabla de Amazon DynamoDB

El formato para una lista de columnas es el siguiente.

```
COPY tablename (column1 [,column2, ...])
```

Si una columna de la tabla de destino se omite de la lista de columnas, COPY carga la expresión DEFAULT de la columna de destino.

Si la columna de destino no tiene un valor predeterminado, COPY intenta cargar NULL.

Si COPY prueba asignar NULL a una columna que está definida como NOT NULL, el comando COPY falla.

Si se incluye una columna IDENTITY en la lista de columnas, también se debe especificar EXPLICIT_IDS. Si se omite una columna IDENTITY, no se puede especificar EXPLICIT_IDS. Si no se especifica una lista de columnas, el comando se comporta como si se especificará una lista de columnas por orden completa, con la omisión de columnas IDENTITY si tampoco se especificó EXPLICIT_IDS.

Si una columna se define con GENERATED BY DEFAULT AS IDENTITY, se puede copiar. Los valores se generan o se actualizan con los que se proporcionan. La opción EXPLICIT_IDS no es obligatoria. COPY no actualiza el límite máximo de la identidad. Para obtener más información, consulte GENERATED BY DEFAULT AS IDENTITY.

Archivo JSONPaths

Cuando se realiza una carga desde archivos de datos en formatos JSON o Avro, COPY asigna de forma automática los elementos de datos en los datos de origen Avro o JSON a las columnas de la tabla de destino. Logra esto al hacer coincidir los nombres de los campos del esquema de Avro con los nombres de las columnas de la lista de columnas o la tabla de destino.

En algunos casos, los nombres de las columnas y los nombres de los campos no coinciden, o bien, es necesario asignarlos a niveles más profundos en la jerarquía de los datos. En estos casos, puede utilizar un archivo JSONPaths para asignar de forma explícita los elementos de datos JSON o Avro a las columnas.

Para obtener más información, consulte Archivo JSONPaths.

Parámetros de formato de datos

De manera predeterminada, el comando COPY espera que los datos de origen estén en texto UTF-8 delimitado por caracteres. El delimitador predeterminado es el carácter de barra vertical ( | ). Si los datos de origen están en otro formato, utilice los siguientes parámetros para especificar el formato de datos:

- FORMAT
- CSV
- DELIMITER
- FIXEDWIDTH
- SHAPEFILE
- AVRO
- JSON
- PARQUET
- ORC

Además de los formatos de datos estándar, COPY admite los siguientes formatos de datos en columnas para COPY de Amazon S3:

- ORC
- PARQUET

Existen ciertas restricciones para poder usar el comando COPY con el formato de columnas. Para obtener más información, consulte Uso de COPY con formatos de datos de columnas.

Parámetros de formato de datos

FORMAT [AS]

(Opcional) Identifica palabras clave de los formatos de datos. Los argumentos FORMAT se describen a continuación.

CSV [ QUOTE [AS] 'quote_character' ]

Permite el uso del formato CSV en los datos de entrada. Para aplicar escape de forma automática a los delimitadores, los caracteres de línea nueva y los retornos de carro, encierre el campo con el carácter especificado por el parámetro QUOTE. El carácter de comilla predeterminado es el

de comilla doble ( " ). Cuando utilice el carácter de comilla dentro del campo, aplique escape al carácter con un carácter de comilla adicional. Por ejemplo, si los caracteres de comillas son comillas dobles, para insertar la cadena `A "quoted" word`, el archivo de entrada debe incluir la cadena `"A ""quoted"" word"`. Cuando se utiliza el parámetro CSV, el delimitador predeterminado es una coma ( , ). Puede especificar un delimitador diferente con el parámetro DELIMITER.

Cuando se encierra un campo entre comillas, se ignora el espacio en blanco entre los delimitadores y los caracteres de comillas. Si el delimitador es un carácter de espacio en blanco, como un tabulador, el delimitador no se trata como un espacio en blanco.

CSV no puede utilizarse con FIXEDWIDTH, REMOVEQUOTES ni ESCAPE.

QUOTE [AS] 'quote_character'

Opcional. Especifica el carácter que se utilizará como carácter de comilla cuando se utilice el parámetro CSV. El carácter predeterminado es una comilla doble ( " ). Si utiliza el parámetro QUOTE para definir un carácter de comilla distinto al de comilla doble, no es necesario que aplique escape a las comillas dobles dentro del campo. El parámetro QUOTE solo puede utilizarse con el parámetro CSV. La palabra clave AS es opcional.

DELIMITER [AS] ['delimiter_char']

Especifica caracteres que se utilizan para separar campos en el archivo de entrada, como un carácter de barra vertical (`|`), una coma (`,`), una tabulación (`\t`) o varios caracteres, como `|~|`. Se admiten los caracteres no imprimibles. Los caracteres también se pueden representar en formato octal como unidades de código UTF-8. Para el formato octal, utilice el formato “\ddd”, donde “d” es un dígito octal (de 0 a 7). El delimitador predeterminado es un carácter de barra vertical (`|`), a menos que se utilice el parámetro CSV, en cuyo caso el delimitador predeterminado es una coma (`,`). La palabra clave AS es opcional. DELIMITER no se puede utilizar con FIXEDWIDTH.

FIXEDWIDTH 'fixedwidth_spec'

Carga los datos de un archivo en el que el ancho de cada columna tiene una longitud fija, en lugar de que las columnas estén separadas por un delimitador. La cadena fixedwidth_spec (especificación_de_ancho_fijo) especifica una etiqueta de columnas definida por el usuario y el ancho de columnas. La etiqueta de columnas puede ser una cadena de texto o un número entero, según lo que decida el usuario. La etiqueta de columnas no tiene relación con el nombre de las columnas. El orden de los pares etiqueta/ancho debe coincidir con el orden de las columnas de la tabla de forma exacta. FIXEDWIDTH no se puede utilizar con CSV ni DELIMITER. En

Amazon Redshift, la longitud de las columnas CHAR y VARCHAR se expresa en bytes, por lo que debe asegurarse de que el ancho de la columna que especifique se adapte a la longitud binaria de caracteres multibyte cuando prepare el archivo que se cargará. Para obtener más información, consulte Tipos de caracteres.

A continuación, se muestra el formato para fixedwidth_spec:

```
'colLabel1:colWidth1,colLabel:colWidth2, ...'
```

SHAPEFILE [ SIMPLIFY [AUTO] ['tolerance'] ]

Permite el uso del formato SHAPEFILE en los datos de entrada. De manera predeterminada, la primera columna del shapefile es una columna `GEOMETRY` o `IDENTITY`. Todas las columnas posteriores tienen el orden especificado en el shapefile.

SHAPEFILE no puede utilizarse con FIXEDWIDTH, REMOVEQUOTES ni ESCAPE.

Para utilizar objetos `GEOGRAPHY` con `COPY FROM SHAPEFILE`, primero captúrelos en una columna `GEOMETRY` y, a continuación, convierta los objetos en objetos `GEOGRAPHY`.

SIMPLIFY [tolerance]

(Opcional) Simplifica toda la geometría durante el proceso de ingesta mediante el algoritmo de Ramer-Douglas-Peucker y la tolerancia dada.

SIMPLIFY AUTO [tolerance]

(Opcional) Simplifica solo la geometría que sea más grandes que el tamaño máximo de geometría. Esta simplificación utiliza el algoritmo de Ramer-Douglas-Peucker y la tolerancia calculada de forma automática si no supera la tolerancia especificada. Este algoritmo calcula el tamaño para almacenar objetos en el marco de la tolerancia especificada. El valor tolerance es opcional.

Para ver ejemplos de carga de shapefiles, consulte Carga de un shapefile en Amazon Redshift.

AVRO [AS] 'avro_option'

Especifica que los datos de origen estén en formato de Avro.

El formato de Avro se admite para COPY de estos servicios y protocolos:

- Amazon S3
- Amazon EMR
- Hosts remotos (SSH)

Avro no es compatible con COPY de DynamoDB.

Avro es un protocolo de serialización de datos. Un archivo de origen Avro incluye un esquema que define la estructura de los datos. El tipo de esquema de Avro debe ser `record`. COPY acepta archivos Avro creados mediante el códec predeterminado sin comprimir como también los códecs de compresión `deflate` y `snappy`. Para obtener más información acerca de Avro, visite Apache Avro.

Los valores válidos para avro_option son los siguientes:

- `'auto'`
- `'auto ignorecase'`
- `'s3://jsonpaths_file'`

El valor predeterminado es `'auto'`.

COPY asigna de forma automática los elementos de datos en los datos de origen Avro a las columnas de la tabla de destino. Logra esto al hacer coincidir los nombres de los campos del esquema de Avro con los nombres de las columnas de la tabla de destino. La coincidencia distingue entre letras mayúsculas y minúsculas en `'auto'`, y no distingue entre letras mayúsculas y minúsculas en `'auto ignorecase'`.

Los nombres de las columnas de las tablas de Amazon Redshift están siempre en minúsculas; por lo tanto, cuando utilice la opción `'auto'`, los nombres de campos coincidentes también deben estar en minúsculas. Si los nombres de los campos no están en minúsculas por completo, puede utilizar la opción `'auto ignorecase'`. Con el argumento predeterminado `'auto'`, COPY solo reconoce el primer nivel de campos o los campos externos de la estructura.

Para asignar explícitamente nombres de columnas a nombres de campos de Avro, puede utilizar Archivo JSONPaths.

De manera predeterminada, COPY intenta hacer coincidir todas las columnas de la tabla de destino con los nombres de campos de Avro. Si lo desea, puede especificar una lista de columnas para cargar un subconjunto de las columnas. Si una columna de la tabla de destino se omite de la lista de columnas, COPY carga la expresión DEFAULT de la columna de destino. Si la columna de destino no tiene un valor predeterminado, COPY intenta cargar NULL. Si se incluye una columna en la lista de columnas y COPY no encuentra un campo coincidente en los datos Avro, COPY intenta cargar NULL en la columna.

Si COPY prueba asignar NULL a una columna que está definida como NOT NULL, el comando COPY falla.

### Esquema de Avro

Un archivo de datos de origen Avro incluye un esquema que define la estructura de los datos. COPY lee el esquema que forma parte del archivo de datos de origen Avro para asignar elementos de datos a las columnas de la tabla de destino. En el siguiente ejemplo, se muestra un esquema de Avro.

```
{
    "name": "person",
    "type": "record",
    "fields": [
        {"name": "id", "type": "int"},
        {"name": "guid", "type": "string"},
        {"name": "name", "type": "string"},
        {"name": "address", "type": "string"}]
}
```

El esquema de Avro se define mediante el formato JSON. El objeto JSON de nivel superior contiene tres pares de nombre-valor con los nombres o las claves: `"name"`, `"type"` y `"fields"`.

Los pares de claves `"fields"` con una matriz de objetos que define el nombre y el tipo de datos de cada campo en la estructura de datos. De manera predeterminada, COPY automáticamente hace coincidir los nombres de los campos con los nombres de las columnas. Los nombres de las columnas están siempre en minúsculas, por lo que los nombres de los campos coincidentes también deben estar en minúsculas, a menos que se especifique la opción `‘auto ignorecase’`. Cualquier nombre de campo que no coincida con el nombre de una columna se ignora. El orden no es importante. En el ejemplo anterior, COPY asigna `id`, `guid`, `name` y `address` a los nombres de columnas.

Con el argumento predeterminado `'auto'`, COPY hace coincidir con las columnas solo los objetos de primer nivel. Para asignar a niveles más profundos en el esquema, o si los nombres de los campos y los de las columnas no coinciden, utilice un archivo JSONPaths para definir el mapeo. Para obtener más información, consulte Archivo JSONPaths.

Si el valor asociado a una clave es un tipo complejo de datos Avro, como byte, matriz, registro, mapeo o enlace, COPY carga el valor como una cadena. Aquí, la cadena es la representación JSON de los datos. COPY carga los tipos de datos enum de Avro como cadenas, donde el contenido es el nombre del tipo. Para ver un ejemplo, consulta COPY de formato JSON.

El tamaño máximo del encabezado del archivo de Avro, incluidos el esquema y los metadatos del archivo, es de 1 MB.

El tamaño máximo de un solo bloque de datos de Avro es de 4 MB. Este es distinto al tamaño máximo de filas. Si se supera el tamaño máximo de un bloque de datos de Avro individual, aunque el tamaño de filas obtenido sea menor al límite de tamaño de filas de 4 MB, el comando COPY falla.

Cuando se calcula el tamaño de las filas, Amazon Redshift cuenta de forma interna los caracteres de barra vertical ( | ) dos veces. Si los datos de entrada contienen un número muy grande de caracteres de barra vertical, es posible que el tamaño de filas supere los 4 MB, aunque el bloque de datos sea menor a 4 MB.

JSON [AS] 'json_option (opción_json)'

Los datos de origen están en formato JSON.

El formato JSON se admite para COPY de estos servicios y protocolos:

- Amazon S3
- COPY de Amazon EMR
- COPY de SSH

JSON no es compatible con COPY de DynamoDB.

Los valores válidos para json_option son los siguientes:

- `'auto'`
- `'auto ignorecase'`
- `'s3://jsonpaths_file'`
- `'noshred'`

El valor predeterminado es `'auto'`. Amazon Redshift no fragmenta los atributos de las estructuras JSON en varias columnas mientras carga un documento JSON.

De manera predeterminada, COPY intenta hacer coincidir todas las columnas de la tabla de destino con las claves de nombres de campos JSON. Si lo desea, puede especificar una lista de columnas para cargar un subconjunto de las columnas. Si las claves de los nombres de campos JSON no están en minúsculas por completo, puede utilizar la opción `'auto ignorecase'` o Archivo JSONPaths para asignar de forma explícita los nombres de las columnas a las claves de los nombres de campos JSON.

Si una columna de la tabla de destino se omite de la lista de columnas, COPY carga la expresión DEFAULT de la columna de destino. Si la columna de destino no tiene un valor predeterminado, COPY intenta cargar NULL. Si se incluye una columna en la lista de columnas y COPY no encuentra un campo coincidente entre los datos JSON, COPY intenta cargar NULL en la columna.

Si COPY prueba asignar NULL a una columna que está definida como NOT NULL, el comando COPY falla.

COPY asigna los elementos de datos en los datos de origen JSON a las columnas de la tabla de destino. Logra esto al hacer coincidir las claves de objetos, o los nombres, en los pares de nombre-valor de origen con los nombres de las columnas de la tabla de destino.

Consulte los siguientes detalles acerca de cada valor json_option:

'auto'

Con esta opción, para la coincidencia, se distingue letras entre mayúsculas y minúsculas. Los nombres de las columnas de las tablas de Amazon Redshift están siempre en minúsculas; por lo tanto, cuando utilice la opción `'auto'`, los nombres de campos JSON coincidentes también deben estar en minúsculas.

'auto ignorecase'

Con esta opción, para la coincidencia, no se distingue letras entre mayúsculas y minúsculas. Los nombres de las columnas de las tablas de Amazon Redshift están siempre en minúsculas; por lo tanto, cuando utilice la opción `'auto ignorecase'`, los nombres de campos JSON correspondientes pueden estar en mayúsculas, en minúsculas o en mayúsculas y minúsculas combinadas.

's3://jsonpaths_file'

Con esta opción, COPY utiliza el archivo JSONPaths mencionado para asignar los elementos de datos en los datos de origen JSON a las columnas de la tabla de destino. El argumento *`s3://jsonpaths_file`* debe ser una clave de objeto de Amazon S3 que referencie de forma explícita un solo archivo. Un ejemplo es `'s3://amzn-s3-demo-bucket/jsonpaths.txt`. El argumento no puede ser un prefijo de clave. Para obtener más información acerca del uso de un archivo JSONPaths, consulte the section called “Archivo JSONPaths”.

En algunos casos, el archivo especificado por `jsonpaths_file` tiene el mismo prefijo que la ruta especificada por `copy_from_s3_objectpath` para los archivos de datos. En tal caso,

COPY lee el archivo JSONPaths como un archivo de datos y devuelve errores. Por ejemplo, supongamos que los archivos de datos utilizan la ruta de objeto `s3://amzn-s3-demo-bucket/my_data.json`, y el archivo JSONPaths es `s3://amzn-s3-demo-bucket/my_data.jsonpaths`. En este caso, COPY intenta cargar `my_data.jsonpaths` como un archivo de datos.

'noshred'

Con esta opción, Amazon Redshift no fragmenta los atributos de las estructuras JSON en varias columnas mientras carga un documento JSON.

Archivo de datos JSON

El archivo de datos JSON contiene un conjunto de objetos o matrices. COPY carga cada objeto o matriz JSON en una fila de la tabla de destino. Cada objeto o matriz correspondiente a una fila debe ser una estructura independiente en el nivel raíz, es decir, no debe ser miembro de otra estructura JSON.

Un objeto JSON comienza y termina con llaves ( { } ), y contiene una colección de pares de nombre-valor desordenada. Cada valor y nombre de los pares están separados por dos puntos y los pares están separados por comas. De manera predeterminada, la clave de objeto, o el nombre, en los pares de nombre-valor debe coincidir con el nombre de la columna correspondiente en la tabla. Los nombres de las columnas de las tablas de Amazon Redshift están siempre en minúsculas; por lo tanto, las claves de los nombres de los campos JSON coincidentes también deben estar en minúsculas. Si los nombres de las columnas y las claves JSON no coinciden, utilice the section called “Archivo JSONPaths” para asignar de forma explícita las columnas a las claves.

El orden de los objetos JSON no importa. Cualquier nombre que no coincida con el nombre de una columna se ignora. A continuación, se muestra la estructura de un objeto JSON simple.

```
{
  "column1": "value1",
  "column2": value2,
  "notacolumn" : "ignore this value"
}
```

Una matriz JSON comienza y termina con corchetes ( [ ] ) y contiene una colección ordenada de valores separados por comas. Si los archivos de datos utilizan matrices, debe especificar un archivo JSONPaths para hacer coincidir los valores con las columnas. A continuación, se muestra la estructura de una matriz JSON simple.

```
["value1", value2]
```

El JSON debe tener un formato correcto. Por ejemplo, los objetos o las matrices no pueden separarse con comas o cualquier otro carácter, a excepción de un espacio en blanco. Las cadenas se deben encerrar entre caracteres de comilla doble. Los caracteres de comillas deben ser las comillas comunes (0x22), no las inclinadas ni las "inteligentes".

El tamaño máximo de un objeto o matriz JSON individual, incluidas las llaves o los corchetes, es de 4 MB. Este es distinto al tamaño máximo de filas. Si se supera el tamaño máximo de un objeto o matriz JSON individual, aunque el tamaño de filas obtenido sea menor al límite de tamaño de filas de 4 MB, el comando COPY falla.

Cuando se calcula el tamaño de las filas, Amazon Redshift cuenta de forma interna los caracteres de barra vertical ( | ) dos veces. Si los datos de entrada contienen un número muy grande de caracteres de barra vertical, es posible que el tamaño de filas supere los 4 MB, aunque el tamaño del objeto sea menor a 4 MB.

COPY carga `\n` como un carácter de línea nueva y `\t` como un tabulador. Para cargar una barra oblicua inversa, aplique escape con una barra oblicua inversa ( `\\` ).

COPY busca el origen de JSON especificado para obtener un objeto o matriz JSON válido con un formato correcto. Si COPY encuentra caracteres que no sean de espacios en blanco antes de localizar una estructura JSON utilizable o entre objetos o matrices JSON válidos, COPY devuelve un error para cada instancia. Estos errores cuentan para el recuento de errores MAXERROR. Cuando el recuento de errores iguala o supera MAXERROR, COPY falla.

Por cada error, Amazon Redshift registra una fila en la tabla de sistema STL_LOAD_ERRORS. La columna LINE_NUMBER registra la última línea del objeto JSON que provocó el error.

Si se especifica IGNOREHEADER, COPY ignora el número de líneas especificado en los datos JSON. Los caracteres de línea nueva en los datos JSON siempre se cuentan para los cálculos IGNOREHEADER.

De manera predeterminada, COPY carga las cadenas vacías como campos vacíos. Si se especifica EMPTYASNULL, COPY carga las cadenas vacías para los campos CHAR y VARCHAR como NULL. Las cadenas vacías para otros tipos de datos, como INT, siempre se cargan con NULL.

Las siguientes opciones no son compatibles con JSON:

- CSV

- DELIMITER
- ESCAPE
- FILLRECORD
- FIXEDWIDTH
- IGNOREBLANKLINES
- NULL AS
- READRATIO
- REMOVEQUOTES

Para obtener más información, consulte COPY de formato JSON. Para obtener más información acerca de las estructuras de datos JSON, visite www.json.org.

Archivo JSONPaths

Si el usuario carga datos de origen Avro o en formato JSON, de manera predeterminada, COPY asigna los elementos de datos de primer nivel en los datos de origen a las columnas de la tabla de destino. Logra esto al hacer coincidir cada nombre, o clave de objeto, en un par de nombre-valor con el nombre de una columna de la tabla de destino.

Si los nombres de las columnas y las claves de objeto no coinciden, o si desea asignar a niveles más profundos en la jerarquía de datos, puede utilizar un archivo JSONPaths para asignar de forma explícita elementos de datos JSON o de Avro a las columnas. El archivo JSONPaths asigna los elementos de datos JSON a las columnas haciendo coincidir el orden de columnas en la lista de columnas o tabla de destino.

El archivo JSONPaths debe contener solo un objeto JSON (no una matriz). El objeto JSON es un par de nombre-valor. La clave de objeto, que es el nombre del par de nombre-valor, debe ser `"jsonpaths"`. El valor del par de nombre-valor es una matriz de expresiones JSONPath. Cada expresión JSONPath hace referencia a un elemento individual en la jerarquía de datos JSON o el esquema de Avro, de forma parecida al modo en que una expresión XPath hace referencia a los elementos de un documento XML. Para obtener más información, consulte Expresiones JSONPath.

Para utilizar un archivo JSONPaths, agregue la palabra clave AVRO o JSON al comando COPY. Especifique el nombre del bucket de S3 y la ruta de objeto del archivo JSONPaths con el siguiente formato.

```
COPY tablename
```

```
FROM 'data_source'
CREDENTIALS 'credentials-args'
FORMAT AS { AVRO | JSON } 's3://jsonpaths_file';
```

El valor `s3://jsonpaths_file` debe ser una clave de objeto de Amazon S3 que referencie de forma explícita un solo archivo, como `'s3://amzn-s3-demo-bucket/jsonpaths.txt'`. No puede ser un prefijo de clave.

En algunos casos, si la carga se realiza desde Amazon S3, el archivo especificado por `jsonpaths_file` tiene el mismo prefijo que la ruta especificada por `copy_from_s3_objectpath` para los archivos de datos. En tal caso, COPY lee el archivo JSONPaths como un archivo de datos y devuelve errores. Por ejemplo, supongamos que los archivos de datos utilizan la ruta de objeto `s3://amzn-s3-demo-bucket/my_data.json`, y el archivo JSONPaths es `s3://amzn-s3-demo-bucket/my_data.jsonpaths`. En este caso, COPY intenta cargar `my_data.jsonpaths` como un archivo de datos.

Si el nombre de clave es una cadena distinta de `"jsonpaths"`, el comando COPY no devuelve un error, pero omite jsonpaths_file y utiliza el argumento `'auto'` en su lugar.

Si sucede alguna de las siguientes situaciones, el comando COPY falla:

- El JSON tiene un formato incorrecto.
- Existe más de un objeto JSON.
- Algún carácter a excepción del espacio en blanco existe fuera del objeto.
- Un elemento de la matriz es una cadena vacía o no es una cadena.

MAXERROR no se aplica al archivo JSONPaths.

El archivo JSONPaths no debe cifrarse aunque la opción ENCRYPTED esté especificada.

Para obtener más información, consulte COPY de formato JSON.

Expresiones JSONPath

El archivo JSONPaths utiliza expresiones JSONPath para asignar los campos de datos a las columnas de destino. Cada expresión JSONPath corresponde a una columna de la tabla de destino de Amazon Redshift. El orden de los elementos de la matriz JSONPath debe coincidir con el orden de las columnas de la tabla de destino o de la lista de columnas, si se utiliza una.

Se requieren caracteres de comillas dobles, como se muestra, para los valores y los nombres de campo. Las comillas deben ser las comillas simples (0x22), no las comillas inclinadas o “inteligentes”.

Si un elemento de objeto indicado por una expresión JSONPath no se encuentra en los datos JSON, COPY intenta cargar un valor NULL. Si el objeto indicado tiene un formato incorrecto, COPY devuelve un error de carga.

Si un elemento de la matriz al que se hace referencia en la expresión JSONPath no se encuentra en los datos de Avro o JSON, COPY registra el siguiente error: `Invalid JSONPath format: Not an array or index out of range.` Elimine de JSONPaths todos los elementos de la matriz que no existan en los datos de origen y verifique que las matrices de los datos de origen tienen el formato correcto.

Las expresiones JSONPath pueden utilizar notaciones con corchetes o con puntos, pero las notaciones no se pueden combinar. En el siguiente ejemplo, se muestran expresiones JSONPath que utilizan la notación con corchetes.

```
{
    "jsonpaths": [
        "$['venuename']",
        "$['venuecity']",
        "$['venuestate']",
        "$['venueseats']"
    ]
}
```

En el siguiente ejemplo, se muestran expresiones JSONPath que utilizan la notación con puntos.

```
{
    "jsonpaths": [
        "$.venuename",
        "$.venuecity",
        "$.venuestate",
        "$.venueseats"
    ]
}
```

En el contexto de la sintaxis de COPY de Amazon Redshift, una expresión JSONPath debe especificar la ruta explícita a un elemento de nombre único en una estructura de datos jerárquica de Avro o JSON. Amazon Redshift no admite ningún elemento de JSONPath, como caracteres comodín

o expresiones de filtro, que pueda resolverse como una ruta ambigua o como varios elementos de nombre.

Para obtener más información, consulte COPY de formato JSON.

Uso de JSONPaths con datos de Avro

En el siguiente ejemplo, se muestra un esquema de Avro con varios niveles.

```
{
    "name": "person",
    "type": "record",
    "fields": [
        {"name": "id", "type": "int"},
        {"name": "guid", "type": "string"},
        {"name": "isActive", "type": "boolean"},
        {"name": "age", "type": "int"},
        {"name": "name", "type": "string"},
        {"name": "address", "type": "string"},
        {"name": "latitude", "type": "double"},
        {"name": "longitude", "type": "double"},
        {
            "name": "tags",
            "type": {
                        "type" : "array",
                        "name" : "inner_tags",
                        "items" : "string"
                    }
        },
        {
            "name": "friends",
            "type": {
                        "type" : "array",
                        "name" : "inner_friends",
                        "items" : {
                                    "name" : "friends_record",
                                    "type" : "record",
                                    "fields" : [
                                                {"name" : "id", "type" : "int"},
                                                {"name" : "name", "type" : "string"}
                                               ]
                                  }
                    }
        },
```

```
        {"name": "randomArrayItem", "type": "string"}
    ]
}
```

En el siguiente ejemplo, se muestra un archivo JSONPaths que utiliza las expresiones AvroPath para hacer referencia al esquema anterior.

```
{
    "jsonpaths": [
        "$.id",
        "$.guid",
        "$.address",
        "$.friends[0].id"
    ]
}
```

El ejemplo de JSONPaths incluye los siguientes elementos:

jsonpaths

El nombre del objeto JSON que contiene las expresiones AvroPath.

[ … ]

Los corchetes encierran la matriz JSON que contiene los elementos de la ruta.

$

El símbolo de dólar hace referencia al elemento raíz en el esquema de Avro, que es la matriz `"fields"`.

"$.id",

El destino de la expresión AvroPath. En esta instancia, el destino es el elemento de la matriz `"fields"` con el nombre `"id"`. Las expresiones están separadas por comas.

"$.friends[0].id"

Los corchetes indican un índice de matriz. Las expresiones JSONPath utilizan una indexación de base cero, por lo que esta expresión hace referencia al primer elemento de la matriz `"friends"` con el nombre `"id"`.

La sintaxis del esquema de Avro requiere el uso de campos internos para definir la estructura de registro y los tipos de datos de matriz. Las expresiones AvroPath ignoran los campos internos. Por

ejemplo, el campo `"friends"` define una matriz denominada `"inner_friends"`, que a su vez define un registro denominado `"friends_record"`. La expresión AvroPath que hace referencia al campo `"id"` puede ignorar los campos adicionales para hacer referencia directamente al campo de destino. Las siguientes expresiones AvroPath hacen referencia a los dos campos que pertenecen a la matriz `"friends"`.

```
"$.friends[0].id"
"$.friends[0].name"
```

Parámetros de formatos de datos en columnas

Además de los formatos de datos estándar, COPY admite los siguientes formatos de datos en columnas para COPY de Amazon S3. Existen ciertas restricciones para poder usar el comando COPY con el formato de columnas. Para obtener más información, consulte Uso de COPY con formatos de datos de columnas.

ORC

Carga los datos desde un archivo que utiliza el formato Optimized Row Columnar (ORC).

PARQUET

Carga los datos desde un archivo que utiliza el formato Parquet.

Parámetros de compresión de archivos

Especifique los siguientes parámetros para cargar desde archivos de datos comprimidos.

Parámetros de compresión de archivos

BZIP2

Un valor que especifica que el o los archivos de entrada están en el formato comprimido bzip2 (archivos .bz2). La operación COPY lee cada archivo comprimido y descomprime los datos a medida que los carga.

GZIP

Un valor que especifica que el o los archivos de entrada están en el formato comprimido gzip (archivos .gz). La operación COPY lee cada archivo comprimido y descomprime los datos a medida que los carga.

LZOP

Un valor que especifica que el o los archivos de entrada están en el formato comprimido lzop (archivos .lzo). La operación COPY lee cada archivo comprimido y descomprime los datos a medida que los carga.

> **Note**
>
> COPY no admite archivos que estén comprimidos con la opción --filter de lzop.

ZSTD

Un valor que especifica que el o los archivos de entrada están en el formato comprimido Zstandard (archivos .zst). La operación COPY lee cada archivo comprimido y descomprime los datos a medida que los carga.

> **Note**
>
> ZSTD solo es compatible con COPY de Amazon S3.

Parámetros de conversión de datos

A medida que carga la tabla, COPY intenta convertir de forma implícita las cadenas de los datos de origen al tipo de datos de la columna de destino. Si necesita especificar una conversión que sea diferente a la del comportamiento predeterminado, o si la conversión predeterminada da lugar a errores, puede administrar las conversiones de datos especificando los siguientes parámetros. Para obtener más información sobre la sintaxis de estos parámetros, consulte Sintaxis de COPY.

- ACCEPTANYDATE
- ACCEPTINVCHARS
- BLANKSASNULL
- DATEFORMAT
- EMPTYASNULL
- ENCODING
- ESCAPE

- EXPLICIT_IDS
- FILLRECORD
- IGNOREBLANKLINES
- IGNOREHEADER
- NULL AS
- REMOVEQUOTES
- ROUNDEC
- TIMEFORMAT
- TRIMBLANKS
- TRUNCATECOLUMNS

Parámetros de conversión de datos

ACCEPTANYDATE

Permite que se cargue cualquier formato de fecha, incluidos los formatos no válidos como `00/00/00 00:00:00`, sin generar un error. Este parámetro aplica solo para las columnas TIMESTAMP y DATE. Siempre utilice ACCEPTANYDATE con el parámetro DATEFORMAT. Si el formato de fecha de los datos no coincide con la especificación DATEFORMAT, Amazon Redshift inserta un valor NULL en ese campo.

ACCEPTINVCHARS [AS] ['replacement_char']

Permite cargar datos en las columnas VARCHAR, aunque los datos contengan caracteres UTF-8 no válidos. Cuando se especifica ACCEPTINVCHARS, COPY sustituye todos los caracteres UTF-8 no válidos por una cadena de la misma longitud que se compone del carácter especificado en replacement_char (carácter_de_sustitución). Por ejemplo, si el carácter de sustitución es "^", un carácter de tres bytes no válido será sustituido por "^^^".

El carácter de sustitución puede ser cualquier carácter ASCII, excepto NULL. El carácter predeterminado es un signo de interrogación (?). Para obtener más información acerca de los caracteres UTF-8 no válidos, consulte Errores de carga de caracteres multibyte.

COPY devuelve el número de filas que contienen caracteres UTF-8 no válidos y agrega una entrada a la tabla de sistema STL_REPLACEMENTS por cada fila afectada, hasta un máximo de 100 filas por cada sector del nodo. También se sustituyen otros caracteres UTF-8 no válidos, pero esos eventos de sustitución no se registran.

Si no se especifica ACCEPTINVCHARS, COPY devuelve un error siempre que encuentra un carácter UTF-8 no válido.

ACCEPTINVCHARS es válido solo para las columnas VARCHAR.

BLANKSASNULL

Carga los campos en blanco, que estén compuestos solo por caracteres de espacios en blanco, como NULL. Esta opción aplica solo para las columnas CHAR y VARCHAR. Los campos en blanco para otros tipos de datos, como INT, siempre se cargan con NULL. Por ejemplo, una cadena que contiene tres caracteres de espacio de forma sucesiva (y ningún otro carácter) se carga como NULL. El comportamiento predeterminado, sin esta opción, es cargar los caracteres de espacio como están.

DATEFORMAT [AS] {'dateformat_string' | 'auto' }

Si no se especifica DATEFORMAT, el formato predeterminado es `'YYYY-MM-DD'`. Por ejemplo, una alternativa de formato válida sería `'MM-DD-YYYY'`.

Si el comando COPY no reconoce el formato de los valores de fecha u hora, o si los valores de fecha u hora usan formatos diferentes, utilice el argumento `'auto'` con el parámetro DATEFORMAT o TIMEFORMAT. El argumento `'auto'` reconoce varios formatos que no se admiten cuando se utiliza una cadena DATEFORMAT y TIMEFORMAT. La palabra clave `'auto'` distingue entre mayúsculas y minúsculas. Para obtener más información, consulte Utilización del reconocimiento automático con DATEFORMAT y TIMEFORMAT.

El formato de fecha puede incluir información de hora (hora, minutos, segundos), pero esta información se ignora. La palabra clave AS es opcional. Para obtener más información, consulte Cadenas TIMEFORMAT y DATEFORMAT.

EMPTYASNULL

Indica que Amazon Redshift debe cargar los campos CHAR y VARCHAR vacíos como NULL. Los campos vacíos para otros tipos de datos, como INT, siempre se cargan con NULL. Los campos vacíos se producen cuando los datos contienen dos delimitadores de forma sucesiva sin caracteres entre los delimitadores. EMPTYASNULL y NULL AS '' (cadena vacía) producen el mismo comportamiento.

ENCODING [AS] file_encoding

Especifica el tipo de codificación de los datos de carga. El comando COPY convierte los datos de la codificación especificada a UTF-8 durante la carga.

Los valores válidos de file_encoding son los siguientes:

- UTF8
- UTF16
- UTF16LE
- UTF16BE
- ISO88591

El valor predeterminado es UTF8.

Los nombres de archivo de origen deben utilizar la codificación UTF-8.

Los siguientes archivos deben utilizar la codificación UTF-8, aunque se especifique una codificación diferente para los datos de carga:

- Archivos de manifiesto
- Archivos JSONPaths

La cadena de argumentos proporcionada con los siguientes parámetros debe utilizar UTF-8:

- FIXEDWIDTH 'fixedwidth_spec'
- ACCEPTINVCHARS 'replacement_char'
- DATEFORMAT 'dateformat_string'
- TIMEFORMAT 'timeformat_string'
- NULL AS 'null_string'

Los archivos de datos de ancho fijo deben utilizar la codificación UTF-8. Los anchos de campo se basan en la cantidad de caracteres, no en el número de bytes.

Todos los datos de carga deben utilizar la codificación especificada. Si COPY encuentra una codificación diferente, omite el archivo y devuelve un error.

Si especifica UTF16, los datos deben tener una marca de orden de bytes (BOM). Si sabe si los datos UTF-16 tienen el formato little-endian (LE) o big-endian (BE), puede utilizar UTF16LE o UTF16BE, independientemente de la presencia de una BOM.

Para utilizar la codificación ISO-8859-1, especifique ISO88591. Para obtener más información, consulte ISO/IEC 8859-1 en Wikipedia.

## ESCAPE

Cuando se especifica este parámetro, el carácter de barra oblicua inversa (\) en los datos de entrada se trata como un carácter de escape. El carácter que sigue inmediatamente al carácter de barra oblicua inversa se carga en la tabla como parte del valor actual de la columna, aunque sea un carácter que por lo general tiene un propósito especial. Por ejemplo, puede utilizar este parámetro para aplicar escape al carácter delimitador, a una comilla, a un carácter de línea nueva incrustado o al carácter de escape en sí cuando cualquiera de estos caracteres es una parte legítima del valor de una columna.

Si especifica el parámetro ESCAPE en combinación con el parámetro REMOVEQUOTES, puede aplicar escape y conservar las comillas (' o '') que, de otro modo, se podrían eliminar. La cadena nula predeterminada, `\N`, funciona como está, pero también se le puede aplicar escape en los datos de entrada como `\\N`. Mientras no especifique una cadena nula alternativa con el parámetro NULL AS, `\N` y `\\N` producirán los mismos resultados.

> **Note**
>
> Al carácter de control `0x00` (NUL) no se le puede aplicar escape y se debe eliminar de los datos de entrada o se debe convertir. Este carácter se trata como un marcador de fin de registro (EOR), lo que provoca que el resto del registro se trunque.

No se puede utilizar el parámetro ESCAPE para las cargas FIXEDWIDTH y no se puede especificar el carácter de escape en sí. El carácter de escape es siempre la barra inversa. También, debe asegurarse de que los datos de entrada contengan el carácter de escape en los lugares adecuados.

Estos son algunos ejemplos de los datos de entrada y los datos cargados que se obtienen cuando se especifica el parámetro ESCAPE. El resultado de la fila 4 supone que también se especifica el parámetro REMOVEQUOTES. Los datos de entrada se componen por dos campos delimitados por barras verticales:

```
1|The quick brown fox\[newline]
jumped over the lazy dog.
2| A\\B\\C
3| A \| B \| C
4| 'A Midsummer Night\'s Dream'
```

Los datos cargados en la columna 2 tendrán este aspecto:

```
The quick brown fox
jumped over the lazy dog.
A\B\C
A|B|C
A Midsummer Night's Dream
```

> **Note**
>
> La aplicación del carácter de escape a los datos de entrada para una carga es responsabilidad del usuario. Existe una excepción a este requisito cuando vuelve a cargar datos que se descargaron anteriormente con el parámetro ESCAPE. En este caso, los datos ya contendrán los caracteres de escape necesarios.

El parámetro ESCAPE no interpreta los códigos octal, hexadecimal, Unicode ni otras notaciones de secuencias de escape. Por ejemplo, si los datos de origen contienen el valor de salto de línea octal (`\012`) e intenta cargar estos datos con el parámetro ESCAPE, Amazon Redshift cargará el valor `012` en la tabla y no interpretará este valor como un salto de línea al que se le está aplicando escape.

Para poder aplicar escape a los caracteres de línea nueva en datos que se originan en plataformas de Microsoft Windows, es posible que necesite utilizar dos caracteres de escape: uno para el retorno de carro y otro para el salto de línea. También puede eliminar los retornos de carro antes de cargar el archivo (por ejemplo, mediante la utilidad dos2unix).

EXPLICIT_IDS

Utilice EXPLICIT_IDS con tablas que tengan columnas IDENTITY si desea sustituir los valores autogenerados con valores explícitos de los archivos de datos de origen para las tablas. Si el comando incluye una lista de columnas, esa lista debe incluir las columnas IDENTITY para utilizar este parámetro. El formato de datos para los valores EXPLICIT_IDS debe coincidir con el formato de IDENTITY especificado por la definición CREATE TABLE.

Cuando ejecute un comando COPY en una tabla con la opción EXPLICIT_IDS, Amazon Redshift no comprueba la singularidad de las columnas IDENTITY de la tabla.

Si una columna se define con GENERATED BY DEFAULT AS IDENTITY, se puede copiar. Los valores se generan o se actualizan con los que se proporcionan. La opción EXPLICIT_IDS no es obligatoria. COPY no actualiza el límite máximo de la identidad.

Para ver un ejemplo de un comando COPY que utiliza EXPLICIT_IDS, consulte Carga de VENUE con valores explícitos para una columna IDENTITY.

FILLRECORD

Permite que se carguen los archivos de datos cuando faltan columnas contiguas al final de alguno de los registros. Las columnas que faltan se cargan como NULL. Para los formatos de texto y CSV, si la columna que falta es una columna VARCHAR, se cargan cadenas de longitud cero en lugar de NULL. Para cargar columnas NULL en columnas VARCHAR mediante texto y CSV, especifique la palabra clave EMPTYASNULL. La sustitución por NULL solo funciona si la definición de la columna permite NULL.

Por ejemplo, si la definición de la tabla contiene cuatro columnas CHAR con valores nulos y un registro contiene los valores `apple, orange, banana, mango`, el comando COPY podría cargar y completar un registro que solo contenga los valores `apple, orange`. Los valores CHAR faltantes se cargarían como valores NULL.

IGNOREBLANKLINES

Ignora las líneas en blanco que solo contienen un salto de línea en un archivo de datos y no prueba cargarlos.

IGNOREHEADER [ AS ] number_rows

Trata los valores number_rows (número_de_filas) especificados como un encabezado de archivo y no los carga. Utilice IGNOREHEADER para omitir encabezados de archivos de todos los archivos en una carga paralela.

NULL AS 'null_string'

Carga los campos en los que el valor de null_string es NULL, donde null_string puede ser cualquier cadena. Si los datos incluyen un carácter nulo, también denominado NUL (UTF-8 0000) o cero binario (0x000), COPY lo trata como cualquier otro carácter. Por ejemplo, un registro que contiene '1' || NUL || '2' se copia como cadena de 3 bytes de longitud. Si un campo solo contiene NUL, puede utilizar NULL AS para reemplazar el carácter nulo con NULL mediante la especificación de `'\0'` o `'\000'`; por ejemplo, `NULL AS '\0'` o `NULL AS '\000'`. Si un campo contiene una cadena que termina con NUL y se especifica NULL AS, la cadena se inserta

con NUL al final. No utilice '\n' (línea nueva) para el valor de null_string. Amazon Redshift reserva '\n' para utilizarlo como delimitador de línea. El valor de null_string predeterminado es '\N'.

> **Note**
>
> Si intenta cargar valores nulos en una columna definida como NOT NULL, el comando COPY producirá un error.

REMOVEQUOTES

Elimina las comillas circundantes de las cadenas en los datos entrantes. Todos los caracteres dentro de las comillas, incluidos los delimitadores, se conservan. Si una cadena tiene una comilla de apertura simple o doble pero no su signo de cierre correspondiente, el comando COPY no puede cargar esa fila y devuelve un error. En la siguiente tabla, se muestran algunos ejemplos simples de cadenas que contienen comillas y los valores cargados que se obtienen.

| Cadena de entrada | Valor cargado con la opción REMOVEQUOTES |
|---|---|
| "The delimiter is a pipe (\|) character" | The delimiter is a pipe (\|) character |
| 'Black' | Black |
| "White" | White |
| Blue' | Blue' |
| Blue' | Value not loaded: error condition (Valor no cargado: condición de error) |
| "Blue | Value not loaded: error condition (Valor no cargado: condición de error) |
| ' ' 'Black' ' ' | ' 'Black' ' |
| ' ' | <white space> |

ROUNDEC

Redondea hacia arriba los valores numéricos cuando la escala del valor de entrada es mayor que la escala de la columna. De manera predeterminada, COPY trunca los valores cuando sea necesario ajustar la escala de la columna. Por ejemplo, si un valor de `20.259` se carga en una columna DECIMAL(8,2), COPY trunca el valor a `20.25` de manera predeterminada. Si se especifica ROUNDEC, COPY redondea el valor a `20.26`. El comando INSERT siempre redondea los valores cuando es necesario que coincidan con la escala de la columna, por lo que un comando COPY con el parámetro ROUNDEC se comporta igual que un comando INSERT.

TIMEFORMAT [AS] {'timeformat_string' | 'auto' | 'epochsecs' | 'epochmillisecs' }

Especifica el formato de hora. Si no se especifica TIMEFORMAT, el formato predeterminado es `YYYY-MM-DD HH:MI:SS` para las columnas TIMESTAMP o `YYYY-MM-DD HH:MI:SSOF` para las columnas TIMESTAMPTZ, donde `OF` es el desplazamiento de la hora universal coordinada (UTC). No puede incluir un especificador de zona horaria en timeformat_string. Para cargar datos TIMESTAMPTZ que estén en un formato diferente al predeterminado, especifique 'auto'. Para obtener más información, consulte Utilización del reconocimiento automático con DATEFORMAT y TIMEFORMAT. Para obtener más información acerca de timeformat_string, consulte Cadenas TIMEFORMAT y DATEFORMAT.

El argumento `'auto'` reconoce varios formatos que no se admiten cuando se utiliza una cadena DATEFORMAT y TIMEFORMAT. Si el comando COPY no reconoce el formato de los valores de fecha u hora, o si los valores de fecha y hora usan formatos diferentes entre sí, utilice el argumento `'auto'` con el parámetro DATEFORMAT o TIMEFORMAT. Para obtener más información, consulte Utilización del reconocimiento automático con DATEFORMAT y TIMEFORMAT.

Si los datos de origen se representan como un tiempo de fecha de inicio, que es el número de segundos o milisegundos desde las 00:00:00 UTC del 1 de enero de 1970, especifique `'epochsecs'` o `'epochmillisecs'`.

Las palabras clave `'auto'`, `'epochsecs'` y `'epochmillisecs'` distinguen entre mayúsculas y minúsculas.

La palabra clave AS es opcional.

TRIMBLANKS

Elimina los caracteres de espacio en blanco del final de una cadena VARCHAR. Este parámetro aplica solo para las columnas con un tipo de datos VARCHAR.

TRUNCATECOLUMNS

Trunca los datos de las columnas al número adecuado de caracteres de modo que se ajuste a la especificación de la columna. Aplica solo para las columnas con un tipo de datos CHAR o VARCHAR y filas de 4 MB de tamaño, o menos.

Operaciones de carga de datos

Administre el comportamiento predeterminado de la operación de carga para solucionar problemas o reducir los tiempos de carga especificando los siguientes parámetros.

- COMPROWS
- COMPUPDATE
- IGNOREALLERRORS
- MAXERROR
- NOLOAD
- STATUPDATE

Parámetros

COMPROWS numrows

Especifica la cantidad de filas que se usarán como el tamaño de muestra para el análisis de compresión. El análisis se ejecuta sobre las filas de cada sector de datos. Por ejemplo, si especifica `COMPROWS 1000000` (1 000 000) y el sistema contiene cuatro sectores totales, se leen y analizan no más de 250 000 filas por cada sector.

Si no se especifica COMPROWS, el tamaño de muestra se establece de manera predeterminada en 100 000 en cada sector. Los valores de COMPROWS inferiores al valor predeterminado de 100 000 filas por cada sector se actualizan automáticamente al valor predeterminado. No obstante, la compresión automática no tendrá lugar si la cantidad de datos que se carga es insuficiente para producir una muestra significativa.

Si el número de COMPROWS es superior a la cantidad de filas del archivo de entrada, el comando COPY avanza y ejecuta el análisis de compresión en todas las filas disponibles. El rango aceptado para este argumento es un número entre 1000 y 2147483647 (2 147 483 647).

COMPUPDATE [PRESET | { ON | TRUE } | { OFF | FALSE } ],

Controla que las codificaciones de compresión se apliquen automáticamente durante un comando COPY.

Cuando COMPUPDATE está preestablecido, el comando COPY selecciona la codificación de compresión para cada columna si la tabla de destino está vacía, incluso si las columnas ya tienen codificaciones distintas de RAW. Actualmente, es posible reemplazar las codificaciones de las columnas especificadas. La codificación de cada columna se basa en el tipo de dato de la columna. No se realiza un muestreo de los datos. Amazon Redshift asigna de forma automática la codificación de compresión de la siguiente manera:

- A las columnas que están definidas como claves de ordenación se les asigna una compresión RAW.
- A las columnas que están definidas como tipos de datos BOOLEAN, REAL o DOUBLE PRECISION se les asigna una compresión RAW.
- Las columnas que se definen como SMALLINT, INTEGER, BIGINT, DECIMAL, DATE, TIMESTAMP o TIMESTAMPTZ tienen asignada la compresión AZ64.
- Las columnas que se definen como CHAR o VARCHAR tienen asignada la compresión LZO.

Cuando se omite COMPUPDATE, el comando COPY elige la codificación de compresión de cada columna solo si la tabla de destino está vacía y no se ha especificado una codificación (distinta de RAW) para ninguna de las columnas. Amazon Redshift determina la codificación de cada columna. No se realiza un muestreo de los datos.

Cuando COMPUPDATE está establecido en ON (o TRUE), o cuando se especifica COMPUDATE sin una opción, el comando COPY aplica la compresión automática si la tabla está vacía, aunque las columnas de la tabla ya tengan codificaciones diferentes a RAW. Actualmente, es posible reemplazar las codificaciones de las columnas especificadas. La codificación de cada columna se basa en un análisis de los datos de muestra. Para obtener más información, consulte Carga de tablas con compresión automática.

Cuando COMPUPDATE está establecido en OFF (o FALSE), se deshabilita la compresión automática. Las codificaciones de las columnas no cambian.

Para obtener información sobre la tabla del sistema para analizar la compresión, consulte STL_ANALYZE_COMPRESSION.

IGNOREALLERRORS

Puede especificar esta opción para ignorar todos los errores que se produzcan durante la operación de carga.

No se puede especificar la opción IGNOREALLERRORS si ya se ha especificado la opción MAXERROR. No se puede especificar la opción IGNOREALLERRORS para formatos de columnas, tales como ORC y Parquet.

MAXERROR [AS] error_count

Si la carga devuelve la cantidad de errores especificada en error_count (número_de_errores) o una cantidad mayor, la carga no se realiza correctamente. Si la carga devuelve menos errores, continúa y devuelve un mensaje INFO que establece el número de filas que no se pudieron cargar. Utilice este parámetro para permitir que las cargas continúen cuando determinadas filas no puedan cargarse en la tabla a causa de errores de formato u otras inconsistencias de los datos.

Establezca este valor en `0` o `1` si desea que la carga falle en el momento que suceda el primer error. La palabra clave AS es opcional. El valor predeterminado de MAXERROR es `0` y el límite `100000`.

El número real de errores notificados puede ser mayor que el MAXERROR especificado debido a la naturaleza paralela de Amazon Redshift. Si algún nodo del clúster de Amazon Redshift detecta que MAXERROR se ha superado, cada nodo notifica todos los errores que ha encontrado.

NOLOAD

Verifica la validez de los archivos de datos sin cargar los datos realmente. Utilice el parámetro NOLOAD para asegurar que los archivos de datos se cargan sin errores antes de ejecutar la carga de datos real. La ejecución de COPY con el parámetro NOLOAD es mucho más rápida que la carga de datos ya que solo analiza los archivos.

STATUPDATE [ { ON | TRUE } | { OFF | FALSE } ]

Controla la actualización y los cálculos automáticos de las estadísticas del optimizador al final de un COPY correcto. De manera predeterminada, cuando el parámetro STATUPDATE no se utiliza, las estadísticas se actualizan de forma automática si inicialmente la tabla está vacía.

Siempre que una inserción de datos en una tabla que no esté vacía cambie el tamaño de la tabla significativamente, le recomendamos actualizar las estadísticas con la ejecución de un comando ANALYZE o mediante el argumento STATUPDATE ON.

Con STATUPDATE ON (o TRUE), las estadísticas se actualizan de forma automática independientemente de si la tabla está vacía inicialmente. Si se utiliza STATUPDATE, el usuario actual debe ser el propietario de la tabla o un super usuario. Si no se especifica STATUPDATE, solo se requiere el permiso INSERT.

Con STATUPDATE OFF (o FALSE), las estadísticas nunca se actualizan.

Para obtener información adicional, consulta Análisis de tablas.

Lista alfabética de parámetros

La siguiente lista proporciona los enlaces a cada descripción de parámetros del comando COPY, ordenadas alfabéticamente.

- ACCEPTANYDATE
- ACCEPTINVCHARS
- the section called “ACCESS_KEY_ID, SECRET_ACCESS_KEY”
- AVRO
- BLANKSASNULL
- BZIP2
- COMPROWS
- COMPUPDATE
- the section called “CREDENTIALS”
- CSV
- DATEFORMAT
- DELIMITER
- EMPTYASNULL
- ENCODING
- ENCRYPTED
- ESCAPE
- EXPLICIT_IDS
- FILLRECORD
- FIXEDWIDTH

- FORMAT
- FROM
- GZIP
- the section called “IAM_ROLE”
- IGNOREALLERRORS
- IGNOREBLANKLINES
- IGNOREHEADER
- JSON
- LZOP
- MANIFEST
- MASTER_SYMMETRIC_KEY
- MAXERROR
- NOLOAD
- NULL AS
- READRATIO
- REGION
- REMOVEQUOTES
- ROUNDEC
- the section called “SESSION_TOKEN”
- SHAPEFILE
- SSH
- STATUPDATE
- TIMEFORMAT
- TRIMBLANKS
- TRUNCATECOLUMNS
- ZSTD

## Notas de uso

### Temas

- Permisos para acceder a otros recursos de AWS

- Uso de COPY con alias de puntos de acceso de Amazon S3
- Carga de datos multibyte desde Amazon S3
- Carga de una columna de tipo de datos GEOMETRY o GEOGRAPHY
- Carga del tipo de datos HLLSKETCH
- Carga de una columna de tipo de datos VARBYTE
- Errores al leer varios archivos
- COPY de formato JSON
- Uso de COPY con formatos de datos de columnas
- Cadenas TIMEFORMAT y DATEFORMAT
- Utilización del reconocimiento automático con DATEFORMAT y TIMEFORMAT

### Permisos para acceder a otros recursos de AWS

Para trasladar datos entre el clúster y otro recurso de AWS, como Amazon S3, Amazon DynamoDB, Amazon EMR o Amazon EC2, el clúster debe tener permiso para acceder al recurso y realizar las acciones necesarias. Por ejemplo, para cargar datos de Amazon S3, COPY debe tener acceso LIST para el bucket y acceso GET para los objetos del bucket. Para obtener información acerca de los permisos mínimos, consulte Permisos de IAM para COPY, UNLOAD y CREATE LIBRARY.

Para obtener la autorización para acceder al recurso, el clúster debe estar autenticado. Puede seleccionar uno de los siguientes métodos de autenticación:

- Control de acceso con base en roles: en el caso del control de acceso basado en roles, se especifica un rol de AWS Identity and Access Management (IAM) que el clúster utiliza para la autenticación y la autorización. Para salvaguardar las credenciales de AWS y la información confidencial, le recomendamos encarecidamente utilizar una autenticación basada en roles.
- Control de acceso basado en claves: en el caso del control de acceso basado en claves, se proporcionan las credenciales de acceso de AWS (ID de clave de acceso y clave de acceso secreta) de un usuario como texto sin formato.

### Control de acceso con base en roles

Con el control de acceso basado en roles, el clúster adopta de forma temporal un rol de IAM en su nombre. Luego, en función de las autorizaciones que se otorgaron al rol, su clúster puede acceder a los recursos de AWS requeridos.

La creación de un rol de IAM es similar a la concesión de permisos a un usuario, en el sentido de que se trata de una identidad de AWS con políticas de permisos que determinan lo que la identidad puede y no puede hacer en AWS. No obstante, en lugar de asociarse exclusivamente a un usuario, cualquier entidad puede asumir un rol cuando lo necesite. Además, un rol no tiene ninguna credencial (una contraseña o claves de acceso) asociada a él. En cambio, si se asocia un rol a un clúster, se crean de forma dinámica claves de acceso y se proporcionan al clúster.

Recomendamos el uso de control de acceso basado en roles porque ofrece un control más preciso y seguro del acceso a los recursos de AWS y a la información confidencial de los usuarios, además de salvaguardar las credenciales de AWS.

La autenticación basada en roles brinda los siguientes beneficios:

- Puede utilizar las herramientas de IAM estándar de AWS para definir un rol de IAM y asociarlo a varios clústeres. Cuando se modifica una política de acceso de un rol, los cambios se aplican automáticamente a todos los clústeres que utilicen ese rol.
- Puede definir políticas de IAM pormenorizadas que concedan permisos para que clústeres y usuarios de bases de datos específicos puedan acceder a determinadas acciones y recursos de AWS.
- El clúster obtiene credenciales de sesión temporales en tiempo de ejecución y actualiza las credenciales según se necesite hasta que se complete la operación. En el caso de que utilice credenciales temporales basadas en claves, si estas vencen antes de completarse la operación, la operación falla.
- El ID de clave de acceso y el ID de clave de acceso secreta no se almacenan ni se transmiten en el código SQL.

Para utilizar el control de acceso basado en roles, primero debe crear un rol de IAM mediante el tipo de función del servicio de Amazon Redshift y, luego, debe adjuntarlo al clúster. Como mínimo, el rol debe tener los permisos especificados en Permisos de IAM para COPY, UNLOAD y CREATE LIBRARY. Para conocer los pasos que se deben realizar para crear un rol de IAM y adjuntarlo a un clúster, consulte Autorización a Amazon Redshift para acceder a otros servicios de AWS en su nombre en la Guía de administración de Amazon Redshift.

Puede agregar un rol a un clúster o ver los roles asociados a un clúster mediante la consola de administración, la CLI o la API de Amazon Redshift. Para obtener más información, consulte Asociación de un rol de IAM a un clúster en la Guía de administración de Amazon Redshift.

Al crear un rol de IAM, IAM devuelve el Amazon Resource Name (ARN, Nombre de recurso de Amazon) del rol. Para especificar un rol de IAM, proporcione el ARN del rol con el parámetro the section called “Uso de IAM_ROLE” o el parámetro the section called “Uso de CREDENTIALS”.

Por ejemplo, suponga que se asocia el siguiente rol al clúster.

```
"IamRoleArn": "arn:aws:iam::0123456789012:role/MyRedshiftRole"
```

En el siguiente ejemplo, el comando COPY utiliza el parámetro IAM_ROLE con el ARN del ejemplo anterior para la autenticación y el acceso a Amazon S3.

```
copy customer from 's3://amzn-s3-demo-bucket/mydata'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole';
```

En el siguiente ejemplo de comando COPY se utiliza el parámetro CREDENTIALS para especificar el rol de IAM.

```
copy customer from 's3://amzn-s3-demo-bucket/mydata'
credentials
'aws_iam_role=arn:aws:iam::0123456789012:role/MyRedshiftRole';
```

Además, un superusuario puede conceder el privilegio ASSUMEROLE a los usuarios y los grupos de bases de datos con objeto de proporcionar acceso a un rol para las operaciones COPY. Para obtener más información, consulta GRANT.

Control de acceso basado en claves

Con el control de acceso basado en claves, se proporciona el ID de clave de acceso y la clave de acceso secreta de un usuario de IAM que está autorizado a acceder a los recursos de AWS que contienen los datos. Puede utilizar los parámetros the section called “Uso de ACCESS_KEY_ID y SECRET_ACCESS_KEY” juntos o el parámetro the section called “Uso de CREDENTIALS”.

> **Note**
>
> Recomendamos encarecidamente que use un rol de IAM para su autenticación, en lugar de brindar un ID de clave de acceso y una clave de acceso secreta como texto sin formato. Si elige el control de acceso basado en claves, nunca utilice las credenciales (raíz) de su cuenta de AWS. Siempre cree un usuario de IAM y proporcione el ID de clave de acceso y la clave

de acceso secreta de ese usuario. Si desea conocer los pasos necesarios para crear un usuario de IAM, consulte Creación de un usuario de IAM en su cuenta de AWS.

Para realizar la autenticación con ACCESS_KEY_ID y SECRET_ACCESS_KEY, sustituya *<access-key-id>* y *<secret-access-key>* por el ID de clave de acceso de un usuario autorizado y la clave de acceso secreta completa, tal y como se muestra a continuación.

```
ACCESS_KEY_ID '<access-key-id>'
SECRET_ACCESS_KEY '<secret-access-key>';
```

Para realizar la autenticación con el parámetro CREDENTIALS, sustituya *<access-key-id>* y *<secret-access-key>* por el ID de clave de acceso de un usuario autorizado y la clave de acceso secreta completa, tal y como se muestra a continuación.

```
CREDENTIALS
'aws_access_key_id=<access-key-id>;aws_secret_access_key=<secret-access-key>';
```

Como mínimo, el usuario de IAM debe tener los permisos que se indican en Permisos de IAM para COPY, UNLOAD y CREATE LIBRARY.

Credenciales de seguridad temporales

Si utiliza el control de acceso basado en claves, puede limitar más el acceso que los usuarios tienen a los datos mediante el uso de credenciales de seguridad temporales. La autenticación basada en roles utiliza de forma automática credenciales temporales.

> **Note**
>
> Le recomendamos encarecidamente utilizar role-based access control en lugar de crear credenciales temporales y proporcionar el ID de clave de acceso y la clave de acceso secreta como texto sin formato. El control de acceso basado en roles utiliza de forma automática credenciales temporales.

Las credenciales de seguridad temporales proporcionan más seguridad debido a su breve vigencia y al hecho de que no se pueden reutilizar cuando vencen. El ID de clave de acceso y la clave de acceso secreta generados con el token no pueden utilizarse sin el token, y un usuario que tenga

estas credenciales de seguridad temporales puede acceder a los recursos solo hasta que estas venzan.

Para conceder a los usuarios acceso temporal a los recursos, debe llamar a las operaciones de API de AWS Security Token Service (AWS STS). Las operaciones de API de AWS STS devuelven credenciales de seguridad temporales compuestas por un token de seguridad, un ID de clave de acceso y una clave de acceso secreta. Emita las credenciales de seguridad temporales a los usuarios que necesiten obtener acceso temporalmente a los recursos. Estos usuarios pueden ser usuarios de IAM existentes o pueden ser usuarios que no utilizan AWS. Para obtener más información acerca de la creación de credenciales de seguridad temporales, consulte Uso de credenciales de seguridad temporales en la Guía del usuario de IAM.

Puede utilizar los parámetros the section called “Uso de ACCESS_KEY_ID y SECRET_ACCESS_KEY” junto con el parámetro the section called “SESSION_TOKEN” o el parámetro the section called “Uso de CREDENTIALS”. También debe suministrar el ID de clave de acceso y la clave de acceso secreta que se proporcionaron con el token.

Para realizar la autenticación con ACCESS_KEY_ID, SECRET_ACCESS_KEY y SESSION_TOKEN, sustituya *<temporary-access-key-id>*, *<temporary-secret-access-key>* y *<temporary-token>*, tal y como se muestra a continuación.

```
ACCESS_KEY_ID '<temporary-access-key-id>'
SECRET_ACCESS_KEY '<temporary-secret-access-key>'
SESSION_TOKEN '<temporary-token>';
```

Para autenticar con CREDENTIALS, incluya `session_token=`*`<temporary-token>`* en la cadena de credenciales, como se muestra a continuación.

```
CREDENTIALS
'aws_access_key_id=<temporary-access-key-id>;aws_secret_access_key=<temporary-secret-
access-key>;session_token=<temporary-token>';
```

En el siguiente ejemplo, se muestra un comando COPY con credenciales de seguridad temporales.

```
copy table-name
from 's3://objectpath'
access_key_id '<temporary-access-key-id>'
secret_access_key '<temporary-secret-access-key>'
session_token '<temporary-token>';
```

En el siguiente ejemplo, se carga la tabla LISTING con credenciales temporales y cifrado de archivos.

```
copy listing
from 's3://amzn-s3-demo-bucket/data/listings_pipe.txt'
access_key_id '<temporary-access-key-id>'
secret_access_key '<temporary-secret-access-key>'
session_token '<temporary-token>'
master_symmetric_key '<root-key>'
encrypted;
```

En el siguiente ejemplo, se carga la tabla LISTING mediante el parámetro CREDENTIALS con credenciales temporales y cifrado de archivos.

```
copy listing
from 's3://amzn-s3-demo-bucket/data/listings_pipe.txt'
credentials
'aws_access_key_id=<temporary-access-key-id>;aws_secret_access_key=<temporary-secret-access-key>;session_token=<temporary-token>;master_symmetric_key=<root-key>'
encrypted;
```

> **Important**
>
> Las credenciales de seguridad temporales deben ser válidas durante toda la operación COPY o UNLOAD. Si las credenciales de seguridad temporales vencen durante la operación, el comando falla y se revierte la transacción. Por ejemplo, si las credenciales de seguridad temporales vencen después de 15 minutos y la operación COPY requiere una hora, el comando COPY falla antes de completarse. Si utiliza un acceso basado en roles, las credenciales de seguridad temporales se actualizan de forma automática hasta que se complete la operación.

### Permisos de IAM para COPY, UNLOAD y CREATE LIBRARY

El rol o usuario de IAM indicado por el parámetro CREDENTIALS debe tener, como mínimo, los siguientes permisos:

- Para utilizar COPY de Amazon S3, debe tener permiso para utilizar LIST sobre el bucket de Amazon S3 y GET sobre los objetos de Amazon S3 que se están cargando y el archivo de manifiesto, si se utiliza alguno.

- Para utilizar COPY de Amazon S3, Amazon EMR y los alojamientos remotos (SSH) con datos en formato JSON deben tener permiso para utilizar LIST y GET con el archivo JSONPaths de Amazon S3, si se utiliza uno.
- Para utilizar COPY de DynamoDB, debe tener permiso para utilizar SCAN y DESCRIBE en la tabla de DynamoDB que se está cargando.
- Para utilizar COPY desde un clúster de Amazon EMR, debe tener permiso para la acción `ListInstances` en el clúster de Amazon EMR.
- Para utilizar UNLOAD en Amazon S3, debe tener permisos para usar GET, LIST y PUT sobre el bucket de Amazon S3 en que se están descargando los archivos de datos.
- Para utilizar CREATE LIBRARY de Amazon S3, debe tener permiso para usar LIST sobre el bucket de Amazon S3 y GET sobre los objetos de Amazon S3 que se están importando.

> **Note**
>
> Si recibe el mensaje de error `S3ServiceException: Access Denied` cuando ejecuta un comando COPY, UNLOAD o CREATE LIBRARY, el clúster no dispone de los permisos de acceso adecuados para Amazon S3.

Puede administrar los permisos de IAM si asocia una política de IAM a un rol de IAM que esté asociado a su clúster, a un usuario o al grupo al que pertenezca su usuario. Por ejemplo, la política administrada `AmazonS3ReadOnlyAccess` concede permisos para utilizar LIST y GET sobre los recursos de Amazon S3. Para obtener más información acerca de las políticas de IAM, consulte Administración de las políticas de IAM en la Guía del usuario de IAM.

Uso de COPY con alias de puntos de acceso de Amazon S3

COPY admite alias de puntos de acceso de Amazon S3. Para obtener más información, consulte Uso de un alias de estilo bucket en su punto de acceso en la Guía del usuario de Amazon Simple Storage Service.

Carga de datos multibyte desde Amazon S3

Si los datos incluyen caracteres multibyte no ASCII (como caracteres cirílicos o chinos) debe cargar los datos en columnas VARCHAR. El tipo de datos VARCHAR admite caracteres UTF-8 de cuatro bytes, pero el tipo de datos CHAR solo acepta caracteres ASCII de un solo byte. No se pueden

cargar caracteres de cinco bytes o más en las tablas de Amazon Redshift. Para obtener más información, consulte Caracteres multibyte.

Carga de una columna de tipo de datos GEOMETRY o GEOGRAPHY

Solo se puede utilizar COPY en columnas `GEOMETRY` o `GEOGRAPHY` a partir de datos de un archivo de texto delimitado por caracteres, como un archivo CSV. Los datos deben tener la forma hexadecimal del formato well-known binary format (ya sea WKB o EWKB) o el formato well-known text format (ya sea WKT o EWKT) y deben caber dentro del tamaño máximo de una sola fila de entrada al comando COPY. Para obtener más información, consulte COPY.

Para obtener información sobre cómo cargar desde un shapefile, consulte Carga de un shapefile en Amazon Redshift.

Para obtener más información acerca del tipo de datos `GEOMETRY` o `GEOGRAPHY`, consulte Consulta de datos espaciales en Amazon Redshift.

Carga del tipo de datos HLLSKETCH

Los bocetos HLL solo se pueden copiar en formato disperso o denso compatible con Amazon Redshift. Para utilizar el comando COPY en bocetos de HyperLogLog, use el formato Base64 para bocetos de HyperLogLog densos y el formato JSON para bocetos de HyperLogLog dispersos. Para obtener más información, consulte Funciones HyperLogLog.

En el siguiente ejemplo, se importan datos de un archivo CSV a una tabla mediante CREATE TABLE y COPY. Primero, el ejemplo crea la tabla `t1` con CREATE TABLE.

```
CREATE TABLE t1 (sketch hllsketch, a bigint);
```

Luego, utiliza COPY para importar datos de un archivo CSV a la tabla `t1`.

```
COPY t1 FROM s3://amzn-s3-demo-bucket/unload/' IAM_ROLE
 'arn:aws:iam::0123456789012:role/MyRedshiftRole' NULL AS 'null' CSV;
```

Carga de una columna de tipo de datos VARBYTE

Puede cargar datos desde un archivo en formato CSV, Parquet y ORC. Para CSV, los datos se cargan desde un archivo en representación hexadecimal de los datos VARBYTE. No se pueden cargar datos VARBYTE con la opción `FIXEDWIDTH`. No se admiten las opciones `ADDQUOTES` ni `REMOVEQUOTES` de COPY. Una columna VARBYTE no se puede utilizar como columna de partición.

### Errores al leer varios archivos

El comando COPY es atómico y de transacciones. En otras palabras, aunque el comando COPY lee datos de varios archivos, todo el proceso se trata como una sola transacción. Si COPY enfrenta un error cuando lee un archivo, vuelve a ejecutarse de forma automática hasta que se agota el tiempo de espera del proceso (consulte statement_timeout) o si los datos no se pueden descargar de Amazon S3 durante un periodo prolongado (entre 15 y 30 minutos), lo que garantiza que cada archivo se cargue solo una vez. Si el comando COPY falla, se anula toda la transacción y se revierten todos los cambios. Para obtener más información acerca de la administración de errores de carga, consulte Solución de problemas en cargas de datos.

Cuando un comando COPY se ha iniciado de forma correcta, no deja de ejecutarse correctamente aunque termine la sesión; por ejemplo, si el cliente se desconecta. No obstante, si el comando COPY está en un bloque de transacción BEGIN ... END que no se completa porque termina la sesión, se revierte toda la transacción, incluida la instrucción COPY. Para obtener más información acerca de las transacciones, consulte BEGIN.

### COPY de formato JSON

La estructura de datos JSON se compone de un conjunto de objetos o matrices. Un objeto JSON comienza y termina con llaves, y contiene una colección desordenada de pares de nombre-valor. Cada nombre y valor está separado por dos puntos y los pares están separados por comas. El nombre es una cadena entre comillas dobles. Los caracteres de comillas deben ser comillas simples (0x22), no las comillas inclinadas o “inteligentes”.

Una matriz JSON comienza y termina con corchetes y contiene una colección ordenada de valores separados por comas. Un valor puede ser una cadena entre comillas dobles, un número, un booleano true o false, nulo, un objeto JSON o una matriz.

Los objetos y las matrices JSON se pueden anidar, lo que habilita una estructura jerárquica de datos. En el siguiente ejemplo, se muestra una estructura de datos JSON con dos objetos válidos.

```
{
    "id": 1006410,
    "title": "Amazon Redshift Database Developer Guide"
}
{
    "id": 100540,
    "name": "Amazon Simple Storage Service User Guide"
}
```

En el siguiente se muestran los mismos datos como dos matrices JSON.

```
[
    1006410,
    "Amazon Redshift Database Developer Guide"
]
[
    100540,
    "Amazon Simple Storage Service User Guide"
]
```

Opciones de COPY para JSON

Cuando utilice COPY con datos en formato JSON, puede especificar las siguientes opciones:

- `'auto'` : COPY carga de forma automática los campos del archivo JSON.
- `'auto ignorecase'`: COPY carga de forma automática los campos del archivo JSON mientras ignora las mayúsculas y las minúsculas de los nombres de los campos.
- `s3://jsonpaths_file`: COPY utiliza un archivo JSONPaths para analizar los datos de origen JSON. Un archivo JSONPaths es un archivo de texto que contiene un solo objeto JSON con el nombre `"jsonpaths"` combinado con una matriz de expresiones JSONPath. Si el nombre es cualquier cadena diferente a `"jsonpaths"`, COPY utiliza el argumento `'auto'` en lugar de utilizar el archivo JSONPaths.

Por obtener ejemplos que muestren cómo cargar datos mediante `'auto'`, `'auto ignorecase'` o un archivo JSONPaths, y mediante matrices y objetos JSON, consulte Ejemplos de Copy de JSON.

Opción JSONPath

En la sintaxis de COPY de Amazon Redshift, una expresión JSONPath especifica la ruta explícita a un elemento de nombre único en una estructura de datos jerárquica JSON, mediante notaciones con corchetes o con puntos. Amazon Redshift no admite ningún elemento de JSONPath, como caracteres comodín o expresiones de filtro, que pueda resolverse como una ruta ambigua o como varios elementos de nombre. Por este motivo, Amazon Redshift no puede analizar estructuras de datos complejas de varios niveles.

El siguiente es un ejemplo de un archivo JSONPaths con expresiones JSONPath que utilizan la notación con corchetes. El símbolo de dólar ($) representa la estructura de nivel raíz.

```
{
```

```
    "jsonpaths": [
       "$['id']",
       "$['store']['book']['title']",
 "$['location'][0]"
    ]
}
```

En el ejemplo anterior, `$['location'][0]` hace referencia al primer elemento de una matriz. JSON utiliza la indexación de matrices de base cero. Los índices de matriz deben ser números enteros positivos (igual o mayor que cero).

En el siguiente ejemplo, se muestra el archivo JSONPaths anterior que utiliza la notación con puntos.

```
{
    "jsonpaths": [
       "$.id",
       "$.store.book.title",
 "$.location[0]"
    ]
}
```

No puede combinar las notaciones con corchetes y con puntos en la matriz `jsonpaths`. Los corchetes se pueden utilizar en la notación con corchetes y en la notación con puntos para hacer referencia a un elemento de matriz.

Cuando se utiliza la notación con puntos, las expresiones JSONPath no pueden contener los siguientes caracteres:

- Comilla simple recta ( ' )
- Punto (.)
- Corchetes ( [ ] ), a menos que se utilice para hacer referencia a una elemento de matriz

Si el valor del par de nombre-valor indicado por una expresión JSONPath es un objeto o una matriz, todo el objeto o la matriz se carga como una cadena, incluidos los corchetes o las llaves. Por ejemplo, suponga que los datos JSON contienen el siguiente objeto.

```
{
    "id": 0,
    "guid": "84512477-fa49-456b-b407-581d0d851c3c",
```

```
    "isActive": true,
    "tags": [
        "nisi",
        "culpa",
        "ad",
        "amet",
        "voluptate",
        "reprehenderit",
        "veniam"
    ],
    "friends": [
        {
            "id": 0,
            "name": "Martha Rivera"
        },
        {
            "id": 1,
            "name": "Renaldo"
        }
    ]
}
```

La expresión JSONPath `$['tags']` luego devuelve el siguiente valor.

```
"["nisi","culpa","ad","amet","voluptate","reprehenderit","veniam"]"
```

La expresión JSONPath `$['friends'][1]` luego devuelve el siguiente valor.

```
"{"id": 1,"name": "Renaldo"}"
```

Cada expresión JSONPath de la matriz `jsonpaths` corresponde a una columna de la tabla de destino de Amazon Redshift. El orden de los elementos de la matriz `jsonpaths` debe coincidir con el orden de las columnas en la tabla de destino o en la lista de columnas, si se utiliza una.

Por obtener ejemplos que muestren cómo cargar datos utilizando un argumento `'auto'` o un archivo JSONPaths y cómo utilizar los objetos y matrices JSON, consulte Ejemplos de Copy de JSON.

Para obtener información sobre cómo copiar varios archivos JSON, consulte Uso de un manifiesto para especificar archivos de datos.

Caracteres de escape en JSON

COPY carga \n como un carácter de línea nueva y \t como un tabulador. Para cargar una barra oblicua inversa, aplique escape con una barra oblicua inversa ( \\ ).

Suponga, por ejemplo, que tiene el siguiente JSON en un archivo denominado escape.json en el bucket s3://amzn-s3-demo-bucket/json/.

```
{
  "backslash": "This is a backslash: \\",
  "newline": "This sentence\n is on two lines.",
  "tab": "This sentence \t contains a tab."
}
```

Ejecute los siguientes comandos para crear la tabla ESCAPES y cargar el JSON.

```
create table escapes (backslash varchar(25), newline varchar(35), tab varchar(35));

copy escapes from 's3://amzn-s3-demo-bucket/json/escape.json'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
format as json 'auto';
```

Consulte la tabla ESCAPES para ver los resultados.

```
select * from escapes;

       backslash        |      newline      |               tab
------------------------+-------------------+----------------------------------
 This is a backslash: \ | This sentence     | This sentence    contains a tab.
                        :  is on two lines.
(1 row)
```

Pérdida de precisión numérica

Puede perder precisión al cargar números de archivos de datos en formato JSON a una columna definida como tipo de datos numéricos. Algunos valores de coma flotante no se representan exactamente igual en distintos sistemas informáticos. Por ello, es posible que los datos que copie de un archivo JSON no se redondeen según lo esperado. Para evitar una pérdida de precisión, le recomendamos que utilice una de las siguientes alternativas:

- Represente el número como una cadena encerrando el valor entre comillas dobles.

- Use ROUNDEC para redondear el número en lugar de truncarlo.
- En vez de utilizar archivos JSON o Avro, use archivos CSV, de ancho fijo o delimitados por caracteres.

### Uso de COPY con formatos de datos de columnas

El comando COPY puede cargar datos de Amazon S3 en los siguientes formatos en columnas:

- ORC
- Parquet

Para ver ejemplos del uso de COPY a partir de formatos de datos en columnas, consulteEjemplos de COPY.

COPY admite los datos que tienen formato de columna con las siguientes consideraciones:

- El bucket de Amazon S3 debe estar en la misma región de AWS que la base de datos de Amazon Redshift.
- Para acceder a los datos de Amazon S3 a través de un punto de conexión de VPC, configure el acceso mediante políticas de IAM y roles de IAM como se describe en Uso de Amazon Redshift Spectrum con el enrutamiento de VPC mejorado en la Guía de administración de Amazon Redshift.
- El comando COPY no aplica automáticamente codificaciones de compresión.
- Solo se admiten los siguientes parámetros de COPY:
  - ACCEPTINVCHARS al copiar desde un archivo ORC o Parquet.
  - FILLRECORD
  - FROM
  - IAM_ROLE
  - CREDENTIALS
  - STATUPDATE
  - MANIFEST
  - EXPLICIT_IDS
- Si COPY encuentra un error al cargar, el comando no se ejecuta correctamente. Los tipos de datos en columnas no aceptan ACCEPTANYDATE y MAXERROR.

- Los mensajes de error se envían al cliente de SQL. Algunos errores se registran en STL_LOAD_ERRORS y STL_ERROR.
- De manera predeterminada, COPY inserta los valores en las columnas de la tabla de destino en el mismo orden en que las columnas aparecen en los archivos de datos con formato de columna. La tabla de destino y el archivo de datos deben tener el mismo número de columnas.
- Si el archivo que especifica para la operación COPY incluye una de las siguientes extensiones, descomprimimos los datos sin necesidad de agregar ningún parámetro:
  - `.gz`
  - `.snappy`
  - `.bz2`
- COPY con los formatos de archivos Parquet y ORC utiliza Redshift Spectrum y el acceso al bucket. Para utilizar COPY para estos formatos, asegúrese de que no haya ninguna política de IAM que bloquee el uso de direcciones URL prefirmadas de Amazon S3. Las URL prefirmadas generadas por Amazon Redshift son válidas durante 1 hora para que Amazon Redshift tenga tiempo suficiente para cargar todos los archivos del bucket de Amazon S3. Se genera una URL prefirmada única para cada archivo escaneado mediante COPY a partir de formatos de datos en columnas. Para las políticas de bucket que incluyen una acción `s3:signatureAge`, asegúrese de establecer el valor en al menos 3 600 000 milisegundos. Para obtener más información, consulte Uso de Amazon Redshift Spectrum con el enrutamiento de VPC mejorado.
- El parámetro REGION no se admite con COPY a partir de formatos de datos en columnas. Aunque su bucket de Amazon S3 y su base de datos estén en la misma Región de AWS, puede producirse un error como, por ejemplo, REGION argument is not supported for PARQUET based COPY.

### Cadenas TIMEFORMAT y DATEFORMAT

El comando COPY utiliza las opciones DATEFORMAT y TIMEFORMAT para analizar los valores de fecha y hora de sus datos de origen. DATEFORMAT y TIMEFORMAT son cadenas formateadas que deben coincidir con el formato de los valores de fecha y hora de sus datos de origen. Por ejemplo, un comando COPY que cargue datos de origen con el valor de fecha `Jan-01-1999` debe incluir la siguiente cadena DATEFORMAT:

```
COPY ...
            DATEFORMAT AS 'MON-DD-YYYY'
```

Para obtener más información sobre cómo administrar las conversiones de datos de COPY, consulte Parámetros de conversión de datos.

Las cadenas DATEFORMAT y TIMEFORMAT pueden contener separadores de fecha y hora (como "`-`", "`/`" o "`:`"), así como los formatos datepart y timepart de la tabla siguiente.

> **Note**
>
> Si no puede hacer coincidir el formato de sus valores de fecha u hora con los siguientes formatos de datepart y timepart, o si tiene valores de fecha y hora que utilizan formatos diferentes entre sí, utilice el argumento `'auto'` con el parámetro DATEFORMAT o TIMEFORMAT. El argumento `'auto'` reconoce varios formatos que no se admiten cuando se utiliza una cadena DATEFORMAT o TIMEFORMAT. Para obtener más información, consulte Utilización del reconocimiento automático con DATEFORMAT y TIMEFORMAT.

| Partes de fecha o de hora | Significado |
|---|---|
| YY | Año sin el siglo |
| AAAA | Año con el siglo |
| MM | Mes como un número |
| MON | Mes con nombre (nombre abreviado o completo) |
| DD | Día del mes como un número |
| HH o HH24 | Hora (formato de 24 horas)<br><br>**Note**<br>En las cadenas con el formato DATETIME de las funciones SQL, HH es lo mismo que HH12. No obstante, en las cadenas DATEFORMAT y TIMEFORMAT para COPY, HH es lo mismo que HH24. |
| HH12 | Hora (formato de 12 horas) |

| Partes de fecha o de hora | Significado |
|---|---|
| MI | Minutos |
| SS | Segundos |
| AM o PM | Indicador meridiano (para formato de 12 horas) |

El formato de fecha predeterminado es YYYY-MM-DD. El formato predeterminado de marca temporal sin zona horaria (TIMESTAMP) es AAAA-MM-DD HH:MI:SS. El formato predeterminado de marca temporal con zona horaria (TIMESTAMPTZ) es AAAA-MM-DD HH:MI:SSOF, donde OF es el desplazamiento de UTC (por ejemplo, - 8:00). No puede incluir un especificador de zona horaria (TZ, tz u OF) en timeformat_string. El campo de segundos (SS) también admite fracciones de segundos hasta un nivel de detalle de microsegundos. Para cargar datos TIMESTAMPTZ que estén en un formato diferente al predeterminado, especifique 'auto'.

A continuación, encontrará algunos ejemplos de fechas u horas que puede encontrar en sus datos de origen, así como las cadenas DATEFORMAT o TIMEFORMAT correspondientes.

| Ejemplo de fecha u hora de los datos de origen | Sintaxis de DATEFORMAT o TIMEFORMAT |
|---|---|
| 03/31/2003 | DATEFORMAT AS 'MM/DD/YYYY' |
| 31 de marzo de 2003 | DATEFORMAT AS 'MON DD, YYYY' |
| 03.31.2003 18:45:05<br><br>03.31.2003 18:45:05.123456 | TIMEFORMAT AS 'MM.DD.YYYY HH:MI:SS' |

Ejemplo

Para ver un ejemplo del uso de TIMEFORMAT, consulte Carga de una marca temporal o de fecha.

### Utilización del reconocimiento automático con DATEFORMAT y TIMEFORMAT

Si especifica `'auto'` como argumento del parámetro DATEFORMAT o TIMEFORMAT, Amazon Redshift reconocerá y convertirá de forma automática el formato de fecha o de hora en los datos de origen. A continuación se muestra un ejemplo.

```
copy favoritemovies from 'dynamodb://ProductCatalog'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
dateformat 'auto';
```

Cuando se utiliza con el argumento `'auto'` para DATEFORMAT y TIMEFORMAT, COPY reconoce y convierte los formatos de fecha y hora enumerados en la tabla en Cadenas TIMEFORMAT y DATEFORMAT. Además, el argumento `'auto'` reconoce los siguientes formatos, que no se admiten cuando se utiliza una cadena DATEFORMAT y TIMEFORMAT.

| Formato | Ejemplo de cadena de entrada válida |
|---|---|
| ISO 8601 | 2019-02-11T05:09:12.195Z |
| Julian | J2451187 |
| BC | Jan-08-95 BC |
| YYYYMMDD HHMISS | 19960108 040809 |
| YYMMDD HHMISS | 960108 040809 |
| YYYY.DDD | 1996.008 |
| YYYY-MM-DD HH:MI:SS.SSS | 1996-01-08 04:05:06.789 |
| DD Mon HH:MI:SS YYYY TZ | 17 dic 07:37:16 1997 PST |
| MM/DD/YYYY HH:MI:SS.SS TZ | 12/17/1997 07:37:16.00 PST |
| YYYY-MM-DD HH:MI:SS+/-TZ | 1997-12-17 07:37:16-08 |

| Formato | Ejemplo de cadena de entrada válida |
| --- | --- |
| DD.MM.YYYY HH:MI:SS TZ | 12.17.1997 07:37:16.00 PST |

El reconocimiento automático no admite epochsecs ni epochmillisecs.

Para probar si un valor de marca temporal o fecha se convertirá automáticamente, utilice una función CAST para intentar convertir la cadena en un valor de marca temporal o fecha. Por ejemplo, los siguientes comandos prueban el valor de marca temporal `'J2345678 04:05:06.789'`:

```
create table formattest (test char(21));
insert into formattest values('J2345678 04:05:06.789');
select test, cast(test as timestamp) as timestamp, cast(test as date) as date from
 formattest;

        test          |      timestamp      | date
----------------------+---------------------+------------
J2345678 04:05:06.789   1710-02-23 04:05:06 1710-02-23
```

Si los datos de origen para una columna DATE incluyen información de la hora, se trunca el componente de tiempo. Si los datos de origen para una columna TIMESTAMP omiten la información de la hora, se usa 00:00:00 para el componente de tiempo.

## Ejemplos de COPY

> **Note**
>
> Estos ejemplos contienen saltos de línea por motivos de legibilidad. No incluya saltos de líneas ni espacios en la cadena credentials-args (credenciales-argumentos).

Temas

- Carga de FAVORITEMOVIES de una tabla de DynamoDB
- Carga de LISTING desde un bucket de Amazon S3
- Carga de LISTING desde un clúster de Amazon EMR
- Uso de un manifiesto para especificar archivos de datos

- Carga de LISTING de un archivo delimitado por la barra vertical (delimitador predeterminado)
- Carga de LISTING utilizando datos en columnas con el formato de Parquet
- Carga de LISTING utilizando datos en columnas con el formato de ORC
- Carga de EVENT con opciones
- Carga de VENUE de un archivo de datos de ancho fijo
- Carga de CATEGORY de un archivo CSV
- Carga de VENUE con valores explícitos para una columna IDENTITY
- Carga de TIME de un archivo GZIP delimitado por la barra vertical
- Carga de una marca temporal o de fecha
- Carga de datos de un archivo con valores predeterminados
- Uso de COPY de datos con la opción ESCAPE
- Ejemplos de Copy de JSON
- Ejemplos de Copy de Avro
- Preparación de archivos para utilizar COPY con la opción ESCAPE
- Carga de un shapefile en Amazon Redshift
- Comando COPY con la opción NOLOAD
- Comando COPY con un delimitador multibyte y la opción ENCODING

Carga de FAVORITEMOVIES de una tabla de DynamoDB

Los AWS SDK incluyen un ejemplo sencillo de la creación de una tabla de DynamoDB denominada Movies (Películas). (Para obtener información sobre este ejemplo, consulte Introducción a DynamoDB). En el siguiente ejemplo, se carga la tabla MOVIES de Amazon Redshift con los datos de la tabla de DynamoDB. La tabla de Amazon Redshift ya debe existir en la base de datos.

```
copy favoritemovies from 'dynamodb://Movies'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
readratio 50;
```

Carga de LISTING desde un bucket de Amazon S3

En el siguiente ejemplo, se carga LISTING desde un bucket de Amazon S3. El comando COPY carga todos los archivos en la carpeta `/data/listing/`.

```
copy listing
from 's3://amzn-s3-demo-bucket/data/listing/'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole';
```

Carga de LISTING desde un clúster de Amazon EMR

En el siguiente ejemplo, se carga la tabla SALES con datos delimitados por tabuladores de archivos comprimidos con lzop en un clúster de Amazon EMR. COPY carga cada archivo en la carpeta `myoutput/` que comienza con `part-`.

```
copy sales
from 'emr://j-SAMPLE2B500FC/myoutput/part-*'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
delimiter '\t' lzop;
```

En el siguiente ejemplo, se carga la tabla SALES con datos en formato JSON en un clúster de Amazon EMR. COPY carga cada archivo en la carpeta `myoutput/json/`.

```
copy sales
from 'emr://j-SAMPLE2B500FC/myoutput/json/'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
JSON 's3://amzn-s3-demo-bucket/jsonpaths.txt';
```

Uso de un manifiesto para especificar archivos de datos

Puede utilizar un manifiesto para asegurarse de que el comando COPY cargue todos los archivos requeridos, y solo esos, de Amazon S3. También puede utilizar un manifiesto cuando necesite cargar varios archivos de diferentes buckets o archivos que no compartan el mismo prefijo.

Por ejemplo, suponga que necesita cargar los siguientes tres archivos: `custdata1.txt`, `custdata2.txt` y `custdata3.txt`. Puede utilizar el siguiente comando para cargar todos los archivos de `amzn-s3-demo-bucket` que comienzan con `custdata` especificando un prefijo:

```
copy category
from 's3://amzn-s3-demo-bucket/custdata'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole';
```

Si debido a un error solo existen dos de los archivos, COPY carga solo esos dos archivos y termina su ejecución correctamente, lo que se traduce en una carga de datos incompleta. Si el bucket

también contiene un archivo no deseado que utiliza el mismo prefijo, como un archivo denominado `custdata.backup`, por ejemplo, COPY también carga ese archivo, lo que se traduce en una carga de datos no deseados.

Para asegurar que se carguen todos los archivos requeridos y evitar que se carguen archivos no deseados, puede utilizar un archivo de manifiesto. El manifiesto es un archivo de texto con formato JSON que enumera los archivos a procesarse por el comando COPY. Por ejemplo, el siguiente manifiesto carga los tres archivos del ejemplo anterior.

```
{
   "entries":[
      {
         "url":"s3://amzn-s3-demo-bucket/custdata.1",
         "mandatory":true
      },
      {
         "url":"s3://amzn-s3-demo-bucket/custdata.2",
         "mandatory":true
      },
      {
         "url":"s3://amzn-s3-demo-bucket/custdata.3",
         "mandatory":true
      }
   ]
}
```

La marca opcional `mandatory` indica si COPY debe terminar en caso de que el archivo no exista. El valor predeterminado es `false`. Independientemente de cualquier configuración obligatoria, COPY termina si no se encuentran archivos. En este ejemplo, COPY devuelve un error si no se encuentra alguno de los archivos. Los archivos no deseados que podrían haberse seleccionado si se especificó solo un prefijo de clave, como `custdata.backup`, se ignoran, ya que no están en el manifiesto.

Cuando la carga se realiza a partir de archivos de datos con formato ORC o Parquet, se necesita un campo `meta`, tal y como se muestra en el siguiente ejemplo.

```
{
   "entries":[
      {
         "url":"s3://amzn-s3-demo-bucket1/orc/2013-10-04-custdata",
         "mandatory":true,
         "meta":{
```

```
            "content_length":99
         }
      },
      {
         "url":"s3://amzn-s3-demo-bucket2/orc/2013-10-05-custdata",
         "mandatory":true,
         "meta":{
            "content_length":99
         }
      }
   ]
}
```

En el siguiente ejemplo, se utiliza un manifiesto llamado `cust.manifest`.

```
copy customer
from 's3://amzn-s3-demo-bucket/cust.manifest'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
format as orc
manifest;
```

Puede usar un manifiesto para cargar archivos de diferentes buckets o archivos que no compartan el mismo prefijo. En el siguiente ejemplo, se muestra el JSON para cargar datos con archivos cuyos nombres comiencen con una marca de fecha.

```
{
  "entries": [
    {"url":"s3://amzn-s3-demo-bucket/2013-10-04-custdata.txt","mandatory":true},
    {"url":"s3://amzn-s3-demo-bucket/2013-10-05-custdata.txt","mandatory":true},
    {"url":"s3://amzn-s3-demo-bucket/2013-10-06-custdata.txt","mandatory":true},
    {"url":"s3://amzn-s3-demo-bucket/2013-10-07-custdata.txt","mandatory":true}
  ]
}
```

El manifiesto puede enumerar los archivos que se encuentran en buckets diferentes, siempre que los buckets estén en la misma región de AWS que el clúster.

```
{
  "entries": [
    {"url":"s3://amzn-s3-demo-bucket1/custdata1.txt","mandatory":false},
    {"url":"s3://amzn-s3-demo-bucket2/custdata1.txt","mandatory":false},
```

```
    {"url":"s3://amzn-s3-demo-bucket2/custdata2.txt","mandatory":false}
  ]
}
```

Carga de LISTING de un archivo delimitado por la barra vertical (delimitador predeterminado)

En el siguiente ejemplo, se muestra un caso muy simple en el que no se especifican opciones y el archivo de entrada contiene el delimitador predeterminado, una barra vertical ("|").

```
copy listing
from 's3://amzn-s3-demo-bucket/data/listings_pipe.txt'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole';
```

Carga de LISTING utilizando datos en columnas con el formato de Parquet

En el siguiente ejemplo, se cargan datos desde una carpeta de Amazon S3 denominada parquet.

```
copy listing
from 's3://amzn-s3-demo-bucket/data/listings/parquet/'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
format as parquet;
```

Carga de LISTING utilizando datos en columnas con el formato de ORC

En el siguiente ejemplo, se cargan datos desde una carpeta de Amazon S3 denominada `orc`.

```
copy listing
from 's3://amzn-s3-demo-bucket/data/listings/orc/'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
format as orc;
```

Carga de EVENT con opciones

En el siguiente ejemplo, se cargan datos delimitados por llaves en la tabla EVENT y se aplican las siguientes reglas:

- Si se utilizan pares de comillas para rodear cualquier cadena de caracteres, estos se eliminan.
- Las cadenas vacías y las que contienen espacios en blanco se cargan como valores NULL.
- La carga falla si se devuelven más de 5 errores.

- Los valores de marca temporal deben cumplir con el formato especificado; por ejemplo, una marca temporal válida es `2008-09-26 05:43:12`.

```
copy event
from 's3://amzn-s3-demo-bucket/data/allevents_pipe.txt'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
removequotes
emptyasnull
blanksasnull
maxerror 5
delimiter '|'
timeformat 'YYYY-MM-DD HH:MI:SS';
```

Carga de VENUE de un archivo de datos de ancho fijo

```
copy venue
from 's3://amzn-s3-demo-bucket/data/venue_fw.txt'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
fixedwidth 'venueid:3,venuename:25,venuecity:12,venuestate:2,venueseats:6';
```

En el ejemplo anterior se supone un archivo de datos con el mismo formato que muestran los datos de ejemplo. En la siguiente muestra, los espacios actúan como marcadores de posición para que todas las columnas tengan el mismo ancho que se indicó en la especificación:

```
1  Toyota Park              Bridgeview  IL0
2  Columbus Crew Stadium    Columbus    OH0
3  RFK Stadium              Washington  DC0
4  CommunityAmerica BallparkKansas City KS0
5  Gillette Stadium         Foxborough  MA68756
```

Carga de CATEGORY de un archivo CSV

Suponga que desea cargar CATEGORY con los valores que se muestran en la siguiente tabla.

| catid | catgroup | catname | catdesc |
|---|---|---|---|
| 12 | Shows | Musicals | Musical theatre |
| 13 | Shows | Plays | All "non-musical" theatre |

| catid | catgroup | catname | catdesc |
|---|---|---|---|
| 14 | Shows | Opera | All opera, light, and "rock" opera |
| 15 | Concerts | Classical | All symphony, concerto, and choir concerts |

En el siguiente ejemplo, se muestra el contenido de un archivo de texto con los valores de campo separados por comas.

```
12,Shows,Musicals,Musical theatre
13,Shows,Plays,All "non-musical" theatre
14,Shows,Opera,All opera, light, and "rock" opera
15,Concerts,Classical,All symphony, concerto, and choir concerts
```

Si carga el archivo con el parámetro DELIMITER para especificar una entrada delimitada por comas, el comando COPY falla debido a que algunos campos de entrada contienen comas. Puede evitar ese problema si utiliza el parámetro CSV y encierra los campos que contienen comas entre comillas. Si los caracteres de comillas aparecen dentro de una cadena entre comillas, deberá aplicar escape duplicando las comillas. El carácter de comilla predeterminado es la comilla doble, por lo que debe aplicar escape a cada comilla doble con una comilla doble adicional. El archivo de entrada nuevo tiene el aspecto siguiente.

```
12,Shows,Musicals,Musical theatre
13,Shows,Plays,"All ""non-musical"" theatre"
14,Shows,Opera,"All opera, light, and ""rock"" opera"
15,Concerts,Classical,"All symphony, concerto, and choir concerts"
```

Si el nombre del archivo es `category_csv.txt`, puede cargarlo con el siguiente comando COPY:

```
copy category
from 's3://amzn-s3-demo-bucket/data/category_csv.txt'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
csv;
```

Además, para evitar la necesidad de aplicar escape a las comillas dobles de la entrada, puede especificar un carácter de comilla diferente con el parámetro QUOTE AS. Por ejemplo, la siguiente versión de `category_csv.txt` utiliza “`%`” como el carácter de comilla.

```
12,Shows,Musicals,Musical theatre
13,Shows,Plays,%All "non-musical" theatre%
14,Shows,Opera,%All opera, light, and "rock" opera%
15,Concerts,Classical,%All symphony, concerto, and choir concerts%
```

El siguiente comando COPY utiliza QUOTE AS para cargar `category_csv.txt`:

```
copy category
from 's3://amzn-s3-demo-bucket/data/category_csv.txt'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
csv quote as '%';
```

Carga de VENUE con valores explícitos para una columna IDENTITY

En el siguiente ejemplo, se supone que en el momento que se creó la tabla VENUE se especificó al menos una columna (como la columna `venueid`) para que sea una columna de IDENTITY. Este comando anula el comportamiento predeterminado de IDENTITY, que genera automáticamente valores para una columna IDENTITY, y en su lugar carga los valores explícitos del archivo venue.txt. Amazon Redshift no comprueba si se cargan valores IDENTITY duplicados en la tabla al utilizar la opción EXLICIT_IDS.

```
copy venue
from 's3://amzn-s3-demo-bucket/data/venue.txt'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
explicit_ids;
```

Carga de TIME de un archivo GZIP delimitado por la barra vertical

En el siguiente ejemplo, se carga la tabla TIME desde un archivo GZIP delimitado por la barra vertical:

```
copy time
from 's3://amzn-s3-demo-bucket/data/timerows.gz'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
gzip
delimiter '|';
```

Carga de una marca temporal o de fecha

En el siguiente ejemplo, se cargan datos con una marca temporal con formato.

> **Note**
>
> El TIMEFORMAT de `HH:MI:SS` también puede admitir fracciones de segundos que sobrepasen SS hasta un nivel de detalle de microsegundos. El archivo `time.txt` utilizado en este ejemplo contiene una fila, `2009-01-12 14:15:57.119568`.

```
copy timestamp1
from 's3://amzn-s3-demo-bucket/data/time.txt'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
timeformat 'YYYY-MM-DD HH:MI:SS';
```

El resultado de este copy es el siguiente:

```
select * from timestamp1;
c1
----------------------------
2009-01-12 14:15:57.119568
(1 row)
```

Carga de datos de un archivo con valores predeterminados

En el siguiente ejemplo, se utiliza una variación de la tabla VENUE de la base de datos TICKIT. Piense en una tabla VENUE_NEW definida con la siguiente instrucción:

```
create table venue_new(
venueid smallint not null,
venuename varchar(100) not null,
venuecity varchar(30),
venuestate char(2),
venueseats integer not null default '1000');
```

Piense en un archivo de datos venue_noseats.txt que no contenga valores para la columna VENUESEATS, como se muestra en el siguiente ejemplo:

```
1|Toyota Park|Bridgeview|IL|
2|Columbus Crew Stadium|Columbus|OH|
3|RFK Stadium|Washington|DC|
4|CommunityAmerica Ballpark|Kansas City|KS|
5|Gillette Stadium|Foxborough|MA|
```

```
6|New York Giants Stadium|East Rutherford|NJ|
7|BMO Field|Toronto|ON|
8|The Home Depot Center|Carson|CA|
9|Dick's Sporting Goods Park|Commerce City|CO|
10|Pizza Hut Park|Frisco|TX|
```

La siguiente instrucción COPY cargará correctamente la tabla del archivo y aplicará el valor DEFAULT ("1000") a la columna omitida:

```
copy venue_new(venueid, venuename, venuecity, venuestate)
from 's3://amzn-s3-demo-bucket/data/venue_noseats.txt'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
delimiter '|';
```

Ahora observe la tabla cargada:

```
select * from venue_new order by venueid;
venueid |         venuename          |    venuecity    | venuestate | venueseats
---------+----------------------------+-----------------+------------+------------
1 | Toyota Park                | Bridgeview      | IL         |       1000
2 | Columbus Crew Stadium      | Columbus        | OH         |       1000
3 | RFK Stadium                | Washington      | DC         |       1000
4 | CommunityAmerica Ballpark  | Kansas City     | KS         |       1000
5 | Gillette Stadium           | Foxborough      | MA         |       1000
6 | New York Giants Stadium    | East Rutherford | NJ         |       1000
7 | BMO Field                  | Toronto         | ON         |       1000
8 | The Home Depot Center      | Carson          | CA         |       1000
9 | Dick's Sporting Goods Park | Commerce City   | CO         |       1000
10 | Pizza Hut Park             | Frisco          | TX         |       1000
(10 rows)
```

Para el siguiente ejemplo, además de suponer que no se incluyen datos de VENUESEATS en el archivo, también se supone que no se incluyen datos de VENUENAME:

```
1||Bridgeview|IL|
2||Columbus|OH|
3||Washington|DC|
4||Kansas City|KS|
5||Foxborough|MA|
6||East Rutherford|NJ|
7||Toronto|ON|
8||Carson|CA|
```

```
9||Commerce City|CO|
10||Frisco|TX|
```

Al usar la misma definición de tabla, la siguiente instrucción COPY falla ya que no se especificó ningún valor DEFAULT para VENUENAME y esta es una columna NOT NULL:

```
copy venue(venueid, venuecity, venuestate)
from 's3://amzn-s3-demo-bucket/data/venue_pipe.txt'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
delimiter '|';
```

Ahora piense en una variación de la tabla VENUE que utilice una columna IDENTITY:

```
create table venue_identity(
venueid int identity(1,1),
venuename varchar(100) not null,
venuecity varchar(30),
venuestate char(2),
venueseats integer not null default '1000');
```

Como en el ejemplo anterior, se supone que la columna VENUESEATS no tiene valores correspondientes en el archivo de origen. La siguiente instrucción COPY carga la tabla correctamente, incluidos los valores de datos IDENTITY predefinidos en lugar de autogenerar esos valores:

```
copy venue(venueid, venuename, venuecity, venuestate)
from 's3://amzn-s3-demo-bucket/data/venue_pipe.txt'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
delimiter '|' explicit_ids;
```

Esta instrucción falla porque no contiene la columna IDENTITY (falta VENUEID en la lista de columnas), pero incluye un parámetro EXPLICIT_IDS:

```
copy venue(venuename, venuecity, venuestate)
from 's3://amzn-s3-demo-bucket/data/venue_pipe.txt'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
delimiter '|' explicit_ids;
```

Esta instrucción falla porque no contiene un parámetro EXPLICIT_IDS:

```
copy venue(venueid, venuename, venuecity, venuestate)
from 's3://amzn-s3-demo-bucket/data/venue_pipe.txt'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
delimiter '|';
```

Uso de COPY de datos con la opción ESCAPE

En el siguiente ejemplo, se muestra cómo cargar caracteres que coincidan con el carácter delimitador (en este caso, la barra vertical). En el archivo de entrada, asegúrese de que a todos los caracteres de barra vertical (|) que desea cargar se les aplicó el carácter de escape de barra oblicua inversa (\). Luego cargue el archivo con el parámetro ESCAPE.

```
$ more redshiftinfo.txt
1|public\|event\|dwuser
2|public\|sales\|dwuser

create table redshiftinfo(infoid int,tableinfo varchar(50));

copy redshiftinfo from 's3://amzn-s3-demo-bucket/data/redshiftinfo.txt'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
delimiter '|' escape;

select * from redshiftinfo order by 1;
infoid |       tableinfo
-------+--------------------
1      | public|event|dwuser
2      | public|sales|dwuser
(2 rows)
```

Sin el parámetro ESCAPE, este comando COPY producirá un error `Extra column(s) found`.

> **Important**
>
> Si carga los datos a través de un comando COPY con el parámetro ESCAPE, también debe especificar el parámetro ESCAPE con el comando UNLOAD para generar el archivo de salida recíproco. De forma similar, si utiliza UNLOAD con el parámetro ESCAPE, necesita usar ESCAPE cuando utilice COPY con los mismos datos.

## Ejemplos de Copy de JSON

En el siguiente ejemplo, cargue la tabla CATEGORY con los siguientes datos.

| CATID | CATGROUP | CATNAME | CATDESC |
|---|---|---|---|
| 1 | Sports | MLB | Major League Baseball |
| 2 | Sports | NHL | National Hockey League |
| 3 | Sports | NFL | National Football League |
| 4 | Sports | NBA | National Basketball Association |
| 5 | Concerts | Classical | All symphony, concerto, and choir concerts |

Temas

- Carga desde datos JSON con la opción 'auto'
- Carga desde datos JSON con la opción 'auto ignorecase'
- Carga desde datos JSON con un archivo JSONPaths
- Carga desde matrices de JSON con un archivo JSONPaths

### Carga desde datos JSON con la opción 'auto'

Para cargar desde datos JSON con la opción `'auto'`, los datos JSON deben estar compuestos por un conjunto de objetos. Los nombres de clave deben coincidir con los nombres de las columnas, pero el orden no importa. A continuación se muestra el contenido de un archivo denominado `category_object_auto.json`.

```
{
    "catdesc": "Major League Baseball",
    "catid": 1,
    "catgroup": "Sports",
    "catname": "MLB"
}
{
    "catgroup": "Sports",
```

```
    "catid": 2,
    "catname": "NHL",
    "catdesc": "National Hockey League"
}
{
    "catid": 3,
    "catname": "NFL",
    "catgroup": "Sports",
    "catdesc": "National Football League"
}
{
    "bogus": "Bogus Sports LLC",
    "catid": 4,
    "catgroup": "Sports",
    "catname": "NBA",
    "catdesc": "National Basketball Association"
}
{
    "catid": 5,
    "catgroup": "Shows",
    "catname": "Musicals",
    "catdesc": "All symphony, concerto, and choir concerts"
}
```

Para cargar contenido desde los archivos de datos JSON del ejemplo anterior, ejecute el siguiente comando COPY.

```
copy category
from 's3://amzn-s3-demo-bucket/category_object_auto.json'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
json 'auto';
```

Carga desde datos JSON con la opción 'auto ignorecase'

Para cargar desde datos JSON con la opción `'auto ignorecase'`, los datos JSON deben estar compuestos por un conjunto de objetos. No es necesario que las mayúsculas y las minúsculas de los nombres de clave coincidan con los nombres de las columnas y el orden no importa. A continuación se muestra el contenido de un archivo denominado `category_object_auto-ignorecase.json`.

```
{
    "CatDesc": "Major League Baseball",
    "CatID": 1,
```

```
    "CatGroup": "Sports",
    "CatName": "MLB"
}
{
    "CatGroup": "Sports",
    "CatID": 2,
    "CatName": "NHL",
    "CatDesc": "National Hockey League"
}
{
    "CatID": 3,
    "CatName": "NFL",
    "CatGroup": "Sports",
    "CatDesc": "National Football League"
}
{
    "bogus": "Bogus Sports LLC",
    "CatID": 4,
    "CatGroup": "Sports",
    "CatName": "NBA",
    "CatDesc": "National Basketball Association"
}
{
    "CatID": 5,
    "CatGroup": "Shows",
    "CatName": "Musicals",
    "CatDesc": "All symphony, concerto, and choir concerts"
}
```

Para cargar contenido desde los archivos de datos JSON del ejemplo anterior, ejecute el siguiente comando COPY.

```
copy category
from 's3://amzn-s3-demo-bucket/category_object_auto ignorecase.json'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
json 'auto ignorecase';
```

Carga desde datos JSON con un archivo JSONPaths

Si los objetos de datos JSON no corresponden directamente con los nombres de las columnas, puede utilizar un archivo JSONPaths para asignar los elementos JSON a las columnas. El orden de los datos de origen JSON no importa, pero el orden de las expresiones del archivo JSONPaths debe

coincidir con el orden de las columnas. Suponga que tiene el siguiente archivo de datos, denominado `category_object_paths.json`.

```
{
    "one": 1,
    "two": "Sports",
    "three": "MLB",
    "four": "Major League Baseball"
}
{
    "three": "NHL",
    "four": "National Hockey League",
    "one": 2,
    "two": "Sports"
}
{
    "two": "Sports",
    "three": "NFL",
    "one": 3,
    "four": "National Football League"
}
{
    "one": 4,
    "two": "Sports",
    "three": "NBA",
    "four": "National Basketball Association"
}
{
    "one": 6,
    "two": "Shows",
    "three": "Musicals",
    "four": "All symphony, concerto, and choir concerts"
}
```

El siguiente archivo JSONPaths, denominado `category_jsonpath.json`, asigna los datos de origen a las columnas de la tabla.

```
{
    "jsonpaths": [
        "$['one']",
        "$['two']",
        "$['three']",
        "$['four']"
```

```
    ]
}
```

Para cargar contenido desde los archivos de datos JSON del ejemplo anterior, ejecute el siguiente comando COPY.

```
copy category
from 's3://amzn-s3-demo-bucket/category_object_paths.json'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
json 's3://amzn-s3-demo-bucket/category_jsonpath.json';
```

Carga desde matrices de JSON con un archivo JSONPaths

Para cargar desde datos JSON que están compuestos por un conjunto de matrices, debe utilizar un archivo JSONPaths para asignar los elementos de las matrices a las columnas. Suponga que tiene el siguiente archivo de datos, denominado `category_array_data.json`.

```
[1,"Sports","MLB","Major League Baseball"]
[2,"Sports","NHL","National Hockey League"]
[3,"Sports","NFL","National Football League"]
[4,"Sports","NBA","National Basketball Association"]
[5,"Concerts","Classical","All symphony, concerto, and choir concerts"]
```

El siguiente archivo JSONPaths, denominado `category_array_jsonpath.json`, asigna los datos de origen a las columnas de la tabla.

```
{
    "jsonpaths": [
        "$[0]",
        "$[1]",
        "$[2]",
        "$[3]"
    ]
}
```

Para cargar contenido desde los archivos de datos JSON del ejemplo anterior, ejecute el siguiente comando COPY.

```
copy category
from 's3://amzn-s3-demo-bucket/category_array_data.json'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
```

```
json 's3://amzn-s3-demo-bucket/category_array_jsonpath.json';
```

### Ejemplos de Copy de Avro

En el siguiente ejemplo, cargue la tabla CATEGORY con los siguientes datos.

| CATID | CATGROUP | CATNAME | CATDESC |
|---|---|---|---|
| 1 | Sports | MLB | Major League Baseball |
| 2 | Sports | NHL | National Hockey League |
| 3 | Sports | NFL | National Football League |
| 4 | Sports | NBA | National Basketball Association |
| 5 | Concerts | Classical | All symphony, concerto, and choir concerts |

Temas

- Carga desde datos de Avro con la opción 'auto'
- Carga desde datos Avro con la opción 'auto ignorecase'
- Carga desde datos de Avro con un archivo JSONPaths

#### Carga desde datos de Avro con la opción 'auto'

Para cargar desde datos de Avro con el argumento `'auto'`, los nombres de campo del esquema de Avro deben coincidir con los nombres de las columnas. Cuando se utiliza el argumento `'auto'`, el orden no importa. A continuación se muestra el esquema para un archivo denominado `category_auto.avro`.

```
{
    "name": "category",
    "type": "record",
    "fields": [
        {"name": "catid", "type": "int"},
        {"name": "catdesc", "type": "string"},
        {"name": "catname", "type": "string"},
        {"name": "catgroup", "type": "string"},
```

```
}
```

Los datos del archivo de Avro están en formato binario, por lo que no son legibles. A continuación se muestra una representación de JSON de los datos del archivo `category_auto.avro`.

```
{
   "catid": 1,
   "catdesc": "Major League Baseball",
   "catname": "MLB",
   "catgroup": "Sports"
}
{
   "catid": 2,
   "catdesc": "National Hockey League",
   "catname": "NHL",
   "catgroup": "Sports"
}
{
   "catid": 3,
   "catdesc": "National Basketball Association",
   "catname": "NBA",
   "catgroup": "Sports"
}
{
   "catid": 4,
   "catdesc": "All symphony, concerto, and choir concerts",
   "catname": "Classical",
   "catgroup": "Concerts"
}
```

Para cargar contenido desde los archivos de datos Avro del ejemplo anterior, ejecute el siguiente comando COPY.

```
copy category
from 's3://amzn-s3-demo-bucket/category_auto.avro'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
format as avro 'auto';
```

Carga desde datos Avro con la opción 'auto ignorecase'

Para cargar contenido desde datos Avro con el argumento `'auto ignorecase'`, no es necesario que las mayúsculas y las minúsculas de los nombres de campo del esquema de Avro coincidan con

las mayúsculas y las minúsculas de los nombres de las columnas. Cuando se utiliza el argumento `'auto ignorecase'`, el orden no importa. A continuación se muestra el esquema para un archivo denominado `category_auto-ignorecase.avro`.

```
{
    "name": "category",
    "type": "record",
    "fields": [
        {"name": "CatID", "type": "int"},
        {"name": "CatDesc", "type": "string"},
        {"name": "CatName", "type": "string"},
        {"name": "CatGroup", "type": "string"},
}
```

Los datos del archivo de Avro están en formato binario, por lo que no son legibles. A continuación se muestra una representación de JSON de los datos del archivo `category_auto-ignorecase.avro`.

```
{
   "CatID": 1,
   "CatDesc": "Major League Baseball",
   "CatName": "MLB",
   "CatGroup": "Sports"
}
{
   "CatID": 2,
   "CatDesc": "National Hockey League",
   "CatName": "NHL",
   "CatGroup": "Sports"
}
{
   "CatID": 3,
   "CatDesc": "National Basketball Association",
   "CatName": "NBA",
   "CatGroup": "Sports"
}
{
   "CatID": 4,
   "CatDesc": "All symphony, concerto, and choir concerts",
   "CatName": "Classical",
   "CatGroup": "Concerts"
}
```

Para cargar contenido desde los archivos de datos Avro del ejemplo anterior, ejecute el siguiente comando COPY.

```
copy category
from 's3://amzn-s3-demo-bucket/category_auto-ignorecase.avro'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
format as avro 'auto ignorecase';
```

Carga desde datos de Avro con un archivo JSONPaths

Si los nombres de campo del esquema de Avro no corresponden directamente con los nombres de las columnas, puede utilizar un archivo JSONPaths para asignar los elementos del esquema a las columnas. El orden de las expresiones del archivo JSONPaths debe coincidir con el orden de las columnas.

Suponga que tiene un archivo de datos denominado `category_paths.avro` que contiene los mismos datos que el ejemplo anterior, pero con el siguiente esquema.

```
{
    "name": "category",
    "type": "record",
    "fields": [
        {"name": "id", "type": "int"},
        {"name": "desc", "type": "string"},
        {"name": "name", "type": "string"},
        {"name": "group", "type": "string"},
        {"name": "region", "type": "string"}
     ]
}
```

El siguiente archivo JSONPaths, denominado `category_path.avropath`, asigna los datos de origen a las columnas de la tabla.

```
{
    "jsonpaths": [
        "$['id']",
        "$['group']",
        "$['name']",
        "$['desc']"
    ]
}
```

Para cargar contenido desde los archivos de datos Avro del ejemplo anterior, ejecute el siguiente comando COPY.

```
copy category
from 's3://amzn-s3-demo-bucket/category_object_paths.avro'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
format avro 's3://amzn-s3-demo-bucket/category_path.avropath ';
```

Preparación de archivos para utilizar COPY con la opción ESCAPE

En el siguiente ejemplo, se describe cómo puede preparar datos para “aplicar escape” a los caracteres de línea nueva antes de importar los datos a una tabla de Amazon Redshift mediante el comando COPY con el parámetro ESCAPE. Sin la preparación de los datos para delimitar los caracteres de línea nueva, Amazon Redshift devuelve errores de carga cuando ejecuta el comando COPY, ya que el carácter de línea nueva, por lo general, se utiliza como un separador de registros.

Por ejemplo, piense en un archivo o una columna de una tabla externa que desea copiar a una tabla de Amazon Redshift. Si el archivo o la columna tiene contenido con formato XML o datos similares, debe asegurarse de que a todos los caracteres de línea nueva (\n) que forman parte del contenido se les aplique escape con la barra oblicua inversa (\).

Un archivo o una tabla que contienen caracteres de línea nueva insertados proporcionan un patrón relativamente fácil para el cual buscar coincidencias. Lo más probable es que cada carácter de línea nueva insertado siga siempre a un carácter > con algunos caracteres de espacio en blanco potenciales (' ' o tabulación) en el medio, como se puede observar en el siguiente ejemplo de un archivo de texto denominado `nlTest1.txt`.

```
$ cat nlTest1.txt
<xml start>
<newline characters provide>
<line breaks at the end of each>
<line in content>
</xml>|1000
<xml>
</xml>|2000
```

Con el siguiente ejemplo, puede ejecutar un procesador de textos para procesar previamente el archivo de origen e insertar caracteres de escape cuando los necesite. (El carácter | está pensado para utilizarse como delimitador para separar datos de columnas cuando se copian en una tabla de Amazon Redshift).

```
$ sed -e ':a;N;$!ba;s/>[[:space:]]*\n/>\\\n/g' nlTest1.txt > nlTest2.txt
```

Del mismo modo, puede utilizar Perl para realizar una operación similar:

```
cat nlTest1.txt | perl -p -e 's/>\s*\n/>\\\n/g' > nlTest2.txt
```

Para adaptar la carga de datos del archivo `nlTest2.txt` en Amazon Redshift, creamos una tabla con dos columnas en dicho servicio. La primera columna c1, es una columna de caracteres que tiene el contenido con formato XML del archivo `nlTest2.txt`. La segunda columna c2 contiene los valores enteros cargados desde el mismo archivo.

Después de ejecutar el comando `sed`, puede cargar correctamente los datos del archivo `nlTest2.txt` en una tabla de Amazon Redshift mediante el parámetro ESCAPE.

> **Note**
>
> Cuando incluya el parámetro ESCAPE con el comando COPY, se aplica escape a un número de caracteres especiales que incluyen la barra oblicua inversa (incluida la línea nueva).

```
copy t2 from 's3://amzn-s3-demo-bucket/data/nlTest2.txt'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
escape
delimiter as '|';

select * from t2 order by 2;

c1           |  c2
-------------+------
<xml start>
<newline characters provide>
<line breaks at the end of each>
<line in content>
</xml>
| 1000
<xml>
</xml>       | 2000
(2 rows)
```

Puede preparar los archivos de datos exportados de bases de datos externas de forma similar. Por ejemplo, con una base de datos de Oracle, puede utilizar la función REPLACE en cada columna afectada de una tabla que desee copiar en Amazon Redshift.

```
SELECT c1, REPLACE(c2, \n',\\n' ) as c2 from my_table_with_xml
```

Además, varias herramientas de extracción, transformación y carga (ETL) y de exportación de bases de datos que generalmente procesan grandes cantidades de datos brindan opciones para especificar caracteres delimitadores y de escape.

Carga de un shapefile en Amazon Redshift

En los siguientes ejemplos, se muestra cómo cargar un ESRI shapefile con COPY. Para obtener más información sobre cómo cargar shapefiles, consulte Carga de un shapefile en Amazon Redshift.

Carga de un shapefile

Los siguientes pasos muestran cómo capturar datos de OpenStreetMap desde Amazon S3 a través del comando COPY. Este ejemplo asume que el archivo shapefile de Noruega del sitio de descarga de Geofabrik se cargó en un bucket de Amazon S3 privado en su región de AWS. Los archivos `.shp`, `.shx` y `.dbf` deben compartir el mismo prefijo y nombre de archivo de Amazon S3.

Ingesta de datos sin simplificación

Los siguientes comandos crean tablas y capturan datos que caben en el tamaño máximo de geometría sin ninguna simplificación. Abra `gis_osm_natural_free_1.shp` en su software de SIG preferido e inspeccione las columnas de esta capa. De manera predeterminada, las columnas IDENTITY o GEOMETRY son las primeras. Cuando una columna GEOMETRY es la primera, puede crear la tabla como se muestra a continuación.

```
CREATE TABLE norway_natural (
   wkb_geometry GEOMETRY,
   osm_id BIGINT,
   code INT,
   fclass VARCHAR,
   name VARCHAR);
```

Cuando una columna IDENTITY es la primera, puede crear la tabla como se muestra a continuación.

```
CREATE TABLE norway_natural_with_id (
```

```
   fid INT IDENTITY(1,1),
   wkb_geometry GEOMETRY,
   osm_id BIGINT,
   code INT,
   fclass VARCHAR,
   name VARCHAR);
```

Ahora, puede capturar los datos mediante COPY.

```
COPY norway_natural FROM 's3://bucket_name/shapefiles/norway/
gis_osm_natural_free_1.shp'
FORMAT SHAPEFILE
CREDENTIALS 'aws_iam_role=arn:aws:iam::123456789012:role/MyRoleName';
INFO: Load into table 'norway_natural' completed, 83891 record(s) loaded successfully
```

También puede capturar los datos como se muestra a continuación.

```
COPY norway_natural_with_id FROM 's3://bucket_name/shapefiles/norway/
gis_osm_natural_free_1.shp'
FORMAT SHAPEFILE
CREDENTIALS 'aws_iam_role=arn:aws:iam::123456789012:role/MyRoleName';
INFO: Load into table 'norway_natural_with_id' completed, 83891 record(s) loaded
 successfully.
```

Ingesta de datos con simplificación

Los siguientes comandos crean una tabla e intentan capturar datos que no caben en el tamaño máximo de la geometría sin ninguna simplificación. Inspeccione el shapefile `gis_osm_water_a_free_1.shp` y cree la tabla adecuada como se muestra a continuación.

```
CREATE TABLE norway_water (
   wkb_geometry GEOMETRY,
   osm_id BIGINT,
   code INT,
   fclass VARCHAR,
   name VARCHAR);
```

Cuando se ejecuta el comando COPY, se produce un error.

```
COPY norway_water FROM 's3://bucket_name/shapefiles/norway/gis_osm_water_a_free_1.shp'
```

```
FORMAT SHAPEFILE
CREDENTIALS 'aws_iam_role=arn:aws:iam::123456789012:role/MyRoleName';
ERROR:  Load into table 'norway_water' failed.  Check 'stl_load_errors' system table
 for details.
```

La consulta de STL_LOAD_ERRORS muestra que la geometría es demasiado grande.

```
SELECT line_number, btrim(colname), btrim(err_reason) FROM stl_load_errors WHERE query
 = pg_last_copy_id();
 line_number |    btrim     |                                  btrim
-------------+--------------
+-------------------------------------------------------------------------
     1184705 | wkb_geometry | Geometry size: 1513736 is larger than maximum supported
 size: 1048447
```

Para superar esto, el parámetro SIMPLIFY AUTO se agrega al comando COPY para simplificar la geometría.

```
COPY norway_water FROM 's3://bucket_name/shapefiles/norway/gis_osm_water_a_free_1.shp'
FORMAT SHAPEFILE
SIMPLIFY AUTO
CREDENTIALS 'aws_iam_role=arn:aws:iam::123456789012:role/MyRoleName';

INFO:  Load into table 'norway_water' completed, 1989196 record(s) loaded successfully.
```

Para ver las filas y las geometrías que se simplificaron, consulte SVL_SPATIAL_SIMPLIFY.

```
SELECT * FROM svl_spatial_simplify WHERE query = pg_last_copy_id();
 query | line_number | maximum_tolerance | initial_size | simplified | final_size |
 final_tolerance
-------+-------------+-------------------+--------------+------------+------------
+----------------------
    20 |     1184704 |                -1 |      1513736 | t          |    1008808 |
 1.276386653895e-05
    20 |     1664115 |                -1 |      1233456 | t          |    1023584 |
 6.11707814796635e-06
```

El uso de SIMPLIFY AUTO max_tolerance con una tolerancia inferior a las calculadas de manera automática probablemente resulte en un error de ingesta. En este caso, utilice MAXERROR para ignorar los errores.

```
COPY norway_water FROM 's3://bucket_name/shapefiles/norway/gis_osm_water_a_free_1.shp'
FORMAT SHAPEFILE
SIMPLIFY AUTO 1.1E-05
MAXERROR 2
CREDENTIALS 'aws_iam_role=arn:aws:iam::123456789012:role/MyRoleName';

INFO:  Load into table 'norway_water' completed, 1989195 record(s) loaded successfully.
INFO:  Load into table 'norway_water' completed, 1 record(s) could not be loaded.
 Check 'stl_load_errors' system table for details.
```

Consulte SVL_SPATIAL_SIMPLIFY de nuevo para identificar el registro que COPY no logró cargar.

```
SELECT * FROM svl_spatial_simplify WHERE query = pg_last_copy_id();
 query | line_number | maximum_tolerance | initial_size | simplified | final_size |
 final_tolerance
-------+-------------+-------------------+--------------+------------+------------
+-----------------
    29 |     1184704 |           1.1e-05 |      1513736 | f          |          0 |
             0
    29 |     1664115 |           1.1e-05 |      1233456 | t          |     794432 |
     1.1e-05
```

En este ejemplo, el primer registro no logró ajustarse; por lo tanto, la columna `simplified` se muestra como false. El segundo registro se cargó en el marco de la tolerancia determinada. No obstante, el tamaño final es mayor que si se utiliza la tolerancia calculada automáticamente sin especificar la tolerancia máxima.

Carga desde un shapefile comprimido

COPY de Amazon Redshift admite la ingesta de datos de un shapefile comprimido. Todos los componentes de shapefile deben tener el mismo prefijo de Amazon S3 y el mismo sufijo de compresión. Por ejemplo, supongamos que desea cargar los datos del ejemplo anterior. En este caso, los archivos `gis_osm_water_a_free_1.shp.gz`, `gis_osm_water_a_free_1.dbf.gz` y `gis_osm_water_a_free_1.shx.gz` deben compartir el directorio de Amazon S3. El comando COPY requiere la opción GZIP, y la cláusula FROM debe especificar el archivo comprimido correcto, como se muestra a continuación.

```
COPY norway_natural FROM 's3://bucket_name/shapefiles/norway/compressed/
gis_osm_natural_free_1.shp.gz'
FORMAT SHAPEFILE
```

```
GZIP
CREDENTIALS 'aws_iam_role=arn:aws:iam::123456789012:role/MyRoleName';
INFO:  Load into table 'norway_natural' completed, 83891 record(s) loaded successfully.
```

Carga de datos en una tabla con un orden de columnas diferente

Si tiene una tabla en la que GEOMETRY no es la primera columna, puede utilizar el mapeo de columnas para asignar columnas a la tabla de destino. Por ejemplo, cree una tabla con osm_id especificado como primera columna.

```
CREATE TABLE norway_natural_order (
   osm_id BIGINT,
   wkb_geometry GEOMETRY,
   code INT,
   fclass VARCHAR,
   name VARCHAR);
```

Luego, capture un shapefile mediante el mapeo de columnas.

```
COPY norway_natural_order(wkb_geometry, osm_id, code, fclass, name)
FROM 's3://bucket_name/shapefiles/norway/gis_osm_natural_free_1.shp'
FORMAT SHAPEFILE
CREDENTIALS 'aws_iam_role=arn:aws:iam::123456789012:role/MyRoleName';
INFO:  Load into table 'norway_natural_order' completed, 83891 record(s) loaded
 successfully.
```

Carga de datos en una tabla con una columna geography

Si tiene una tabla con una columna GEOGRAPHY, primero captúrela en una columna GEOMETRY y, a continuación, convierta los objetos en objetos GEOGRAPHY. Por ejemplo, después de copiar el shapefile en una columna GEOMETRY, modifique la tabla para agregar una columna de tipo de datos GEOGRAPHY.

```
ALTER TABLE norway_natural ADD COLUMN wkb_geography GEOGRAPHY;
```

A continuación, convierta las columnas geometry en geography.

```
UPDATE norway_natural SET wkb_geography = wkb_geometry::geography;
```

De forma opcional, puede eliminar la columna GEOMETRY.

```
ALTER TABLE norway_natural DROP COLUMN wkb_geometry;
```

Comando COPY con la opción NOLOAD

Para validar los archivos de datos antes de que los datos se carguen, use la opción NOLOAD con el comando COPY. Amazon Redshift analiza el archivo de entrada y muestra los errores que se producen. En el siguiente ejemplo se utiliza la opción NOLOAD donde no se carga ninguna fila en la tabla.

```
COPY public.zipcode1
FROM 's3://amzn-s3-demo-bucket/mydata/zipcode.csv'
DELIMITER ';'
IGNOREHEADER 1 REGION 'us-east-1'
NOLOAD
CREDENTIALS 'aws_iam_role=arn:aws:iam::123456789012:role/myRedshiftRole';

Warnings:
Load into table 'zipcode1' completed, 0 record(s) loaded successfully.
```

Comando COPY con un delimitador multibyte y la opción ENCODING

En el siguiente ejemplo, se carga LATIN1 desde un archivo de Amazon S3 que contiene datos multibyte. El comando COPY especifica el delimitador en formato octal \302\246\303\254 para separar los campos del archivo de entrada, que está codificado como ISO-8859-1. Para especificar el mismo delimitador en UTF-8, indique DELIMITER '¦ì'.

```
COPY latin1
FROM 's3://amzn-s3-demo-bucket/multibyte/myfile'
IAM_ROLE 'arn:aws:iam::123456789012:role/myRedshiftRole'
DELIMITER '\302\246\303\254'
ENCODING ISO88591
```

# CREATE DATABASE

Crea una nueva base de datos.

Para crear una base de datos, debe ser un superusuario o tener el privilegio CREATEDB. Para crear una base de datos asociada a una integración sin ETL, debe ser un superusuario o tener los privilegios CREATEDB y CREATEUSER.

No se puede ejecutar CREATE DATABASE en un bloque de transacción (BEGIN ... END). Para obtener más información acerca de las transacciones, consulte Aislamiento serializable.

## Sintaxis

```
CREATE DATABASE database_name
[ { [
      FROM INTEGRATION '<integration_id>'[ DATABASE '<source_database>' ]
      [ SET ]
      [ ACCEPTINVCHARS [=] { TRUE | FALSE }]
      [ QUERY_ALL_STATES [=] { TRUE | FALSE }]
      [ REFRESH_INTERVAL <interval> ]
      [ TRUNCATECOLUMNS [=] { TRUE | FALSE } ]
      [ HISTORY_MODE [=] {TRUE | FALSE} ]
    ]
    [ WITH ]
    [ OWNER [=] db_owner ]
    [ CONNECTION LIMIT { limit | UNLIMITED } ]
    [ COLLATE { CASE_SENSITIVE | CASE_INSENSITIVE } ]
    [ ISOLATION LEVEL { SERIALIZABLE | SNAPSHOT } ]
  }
  | { FROM { { ARN '<arn>' } { WITH DATA CATALOG SCHEMA '<schema>' | WITH NO DATA
 CATALOG SCHEMA } } }
  | { IAM_ROLE  {default | 'SESSION' | 'arn:aws:iam::<account-id>:role/<role-name>' } }
  | { [ WITH PERMISSIONS ] FROM DATASHARE datashare_name OF [ ACCOUNT account_id ]
 NAMESPACE namespace_guid }
]
```

## Parámetros

database_name

Nombre de la nueva base de datos. Para obtener más información acerca de los nombres válidos, consulte Nombres e identificadores.

FROM INTEGRATION '<id_de_integración>' [ DATABASE '<base_de_datos_de_origen>' ]

Especifica si se debe crear la base de datos mediante un identificador de integración sin ETL. Puede recuperar `integration_id` de la vista del sistema SVV_INTEGRATION. Para

las integraciones sin ETL de Aurora PostgreSQL, también debe especificar el nombre de `source_database`, que también se puede recuperar de SVV_INTEGRATION.

Para ver un ejemplo, consulta Creación de bases de datos para recibir los resultados de las integraciones sin ETL. Para obtener más información sobre la creación de bases de datos con integraciones sin ETL, consulte Creación de bases de datos de destino en Amazon Redshift en la Guía de administración de Amazon Redshift.

SET

Palabra clave opcional.

ACCEPTINVCHARS [=] { TRUE | FALSE }

La cláusula ACCEPTINVCHARS establece si las tablas de integración sin ETL continúan con la ingesta cuando se detectan caracteres no válidos para el tipo de datos VARCHAR. Cuando se encuentran caracteres no válidos, se reemplazan por un carácter ? predeterminado.

QUERY_ALL_STATES [=] { TRUE | FALSE }

La cláusula QUERY_ALL_STATES establece si las tablas de integración sin ETL pueden consultarse en todos los estados (`Synced`, `Failed`, `ResyncRequired` y `ResyncInitiated`). De forma predeterminada, una tabla de integración sin ETL solo puede consultarse en estado `Synced`.

REFRESH_INTERVAL <intervalo>

La cláusula REFRESH_INTERVAL establece el intervalo de tiempo aproximado, en segundos, para actualizar los datos desde el origen sin ETL en la base de datos de destino. El valor se puede establecer de 0 a 432 000 segundos (5 días) para las integraciones sin ETL cuyo tipo de origen sea Aurora MySQL, Aurora PostgreSQL o RDS para MySQL. Para las integraciones sin ETL de Amazon DynamoDB, el valor se puede establecer entre 900 y 432 000 segundos (15 minutos y 5 días). El `interval` predeterminado es cero (0) segundos para las integraciones sin ETL cuyo tipo de origen sea Aurora MySQL, Aurora PostgreSQL o RDS para MySQL. Para las integraciones sin ETL de Amazon DynamoDB, el `interval` predeterminado es de 900 segundos (15 minutos).

TRUNCATECOLUMNS [=] { TRUE | FALSE }

La cláusula TRUNCATECOLUMNS establece si las tablas de integración sin ETL continúan con la ingesta cuando los valores de la columna VARCHAR o los atributos de columna SUPER exceden el límite. Cuando TRUE, los valores se truncan para que quepan en la columna y los valores de los atributos JSON desbordados se truncan para que quepan en la columna SUPER.

HISTORY_MODE [=] {TRUE | FALSE}

Cláusula que especifica si Amazon Redshift establecerá el modo de historial para todas las tablas nuevas de la base de datos especificada. Esta opción solo se aplica a las bases de datos creadas para la integración sin ETL.

La cláusula HISTORY_MODE puede establecerse en `TRUE` o `FALSE`. El valor predeterminado es `FALSE`. Para obtener información sobre HISTORY_MODE, consulte Modo de historial en la Guía de administración de Amazon Redshift.

WITH

Palabra clave opcional.

OWNER [=] db_owner

Especifica el nombre de usuario del propietario de la base de datos.

CONNECTION LIMIT { limit | UNLIMITED }

La cantidad máxima de conexiones a la base de datos que los usuarios pueden tener abiertas al mismo tiempo. Este límite no se aplica a los superusuarios. Use la palabra clave UNLIMITED para permitir la cantidad máxima de conexiones simultáneas. También puede aplicarse un límite de la cantidad de conexiones de cada usuario. Para obtener más información, consulte CREAR USUARIO. El valor predeterminado es UNLIMITED. Para ver las conexiones actuales, consulte la vista del sistema STV_SESSIONS.

> **Note**
>
> Si se aplican los límites de conexión tanto para usuarios como para bases de datos, debe haber una ranura de conexión sin utilizar disponible dentro de ambos límites cuando un usuario intenta conectarse.

COLLATE { CASE_SENSITIVE | CASE_INSENSITIVE }

Se trata de una cláusula que especifica si la búsqueda o la comparación de cadenas es CASE_SENSITIVE o CASE_INSENSITIVE. El valor predeterminado es CASE_SENSITIVE.

No se admite la intercalación cuando se crea una base de datos a partir de un recurso compartido de datos.

ISOLATION LEVEL { SERIALIZABLE | SNAPSHOT }

Una cláusula que especifica el nivel de aislamiento utilizado cuando las consultas se ejecutan en una base de datos.

- Aislamiento SERIALIZABLE: proporciona serialización completa para transacciones simultáneas. Para obtener más información, consulte Aislamiento serializable.
- Aislamiento SNAPSHOT: proporciona un nivel de aislamiento con protección contra conflictos de actualización y eliminación. Es el valor predeterminado para una base de datos creada en un clúster aprovisionado o espacio de nombres sin servidor.

Puede ver qué modelo de simultaneidad ejecuta la base de datos de la siguiente manera:

- Consulte la vista de catálogo STV_DB_ISOLATION_LEVEL. Para obtener más información, consulte STV_DB_ISOLATION_LEVEL.

```
SELECT * FROM stv_db_isolation_level;
```

- Consulte la vista PG_DATABASE_INFO.

```
SELECT datname, datconfig FROM pg_database_info;
```

  El nivel de aislamiento por base de datos aparece junto a la clave `concurrency_model`. Un valor de 1 indica SNAPSHOT. Un valor de 2 denota SERIALIZABLE.

En las bases de datos de Amazon Redshift, el aislamiento SERIALIZABLE y SNAPSHOT son tipos de niveles de aislamiento serializable. Es decir, las lecturas sucias, las lecturas no repetibles y las lecturas fantasma se impiden de acuerdo con el estándar SQL. Los niveles de aislamiento garantizan que una transacción opera en una instantánea de datos en el momento en el que aparece cuando comienza la transacción y que ninguna otra transacción puede cambiar esa instantánea. Sin embargo, el aislamiento SNAPSHOT no proporciona serialización completa porque no impide las inserciones ni actualizaciones de sesgos de escritura en diferentes filas de tablas.

En el siguiente escenario se ilustran actualizaciones de sesgos de escritura en las que se utiliza el nivel de aislamiento SNAPSHOT. Una tabla llamada `Numbers` contiene una columna denominada `digits` que contiene los valores `0` y `1`. La instrucción UPDATE de cada usuario no se superpone a la del otro usuario. Sin embargo, los valores `0` y `1` se intercambian. El SQL que ejecutan sigue esta escala de tiempo con los siguientes resultados:

| Tiempo | Acción del usuario | Acción del usuario 2 |
|---|---|---|
| 1 | BEGIN; | |
| 2 | | BEGIN; |
| 3 | SELECT * FROM Numbers<br>`digits`<br>`-`<br>`-----`<br>`0`<br>`1` | |
| 4 | | SELECT * FROM Numbers;<br>`digits`<br>`------`<br>`0`<br>`1` |
| 5 | UPDATE Numbers SET digits=0 WHERE digits=1; | |
| 6 | SELECT * FROM Numbers | |

| Tiempo | Acción del usuario | Acción del usuario 2 |
|---|---|---|
| | `digits`<br>`-`<br>`-----`<br>`0`<br>`0` | |
| 7 | COMMIT | |
| 8 | | Actualizar número SET dígitos=0 WHERE dígitos=1; |
| 9 | | SELECT * FROM Numbers;<br>`digits`<br>`------`<br>`1`<br>`1` |
| 10 | | COMMIT; |
| 11 | SELECT * FROM Numbers<br>`digits`<br>`-`<br>`-----`<br>`1`<br>`0` | |

| Tiempo | Acción del usuario | Acción del usuario 2 |
| --- | --- | --- |
| 12 | | SELECT * FROM Numbers;<br><br>digits<br>------<br>1<br>0 |

Si se ejecuta el mismo escenario mediante aislamiento serializable, Amazon Redshift termina el usuario 2 debido a una infracción serializable y devuelve un error 1023. Para obtener más información, consulte Cómo corregir errores de aislamiento serializable. En este caso, solo el usuario 1 puede confirmar correctamente. No todas las cargas de trabajo requieren aislamiento serializable como requisito, en cuyo caso el aislamiento de instantáneas es suficiente como nivel de aislamiento de destino para la base de datos.

FROM ARN '<ARN>'

ARN de base de datos de AWS Glue que se puede utilizar para crear la base de datos.

{ WITH DATA CATALOG SCHEMA '<esquema>' | WITH NO DATA CATALOG SCHEMA }

> **Note**
>
> Este parámetro solo se puede aplicar si el comando CREATE DATABASE también utiliza el parámetro FROM ARN.

Especifica si se va a crear la base de datos mediante un esquema para acceder a los objetos de AWS Glue Data Catalog.

IAM_ROLE { default | 'SESSION' | 'arn:aws:iam::*<Cuenta de AWS-id>*:role/*<role-name>*' }

> **Note**
>
> Este parámetro solo se puede aplicar si el comando CREATE DATABASE también utiliza el parámetro FROM ARN.

Si especifica un rol de IAM asociado al clúster al ejecutar el comando CREATE DATABASE, Amazon Redshift utilizará las credenciales del rol cuando ejecute consultas en la base de datos.

Especificar la palabra clave `default` significa utilizar el rol de IAM establecido como predeterminado y asociado al clúster.

Use `'SESSION'` si se conecta al clúster de Amazon Redshift mediante una identidad federada y acceda a las tablas desde el esquema externo creado con este comando. Para ver un ejemplo de utilización de una identidad federada, consulte Uso de una identidad federada para administrar el acceso de Amazon Redshift a los recursos locales y a las tablas externas de Amazon Redshift Spectrum, lo que explica cómo configurar la identidad federada.

Utilice el nombre de recurso de Amazon (ARN), de un rol de IAM que el clúster utiliza para la autenticación y la autorización. Como mínimo, el rol de IAM debe tener permiso para realizar una operación LIST en el bucket de Amazon S3 al que se accederá y una operación GET en los objetos de Amazon S3 que el bucket contiene. Para obtener más información sobre cómo utilizar IAM_ROLE al crear una base de datos que utilice AWS Glue Data Catalog para los recursos compartidos de datos, consulte Uso de recursos compartidos de datos administrados por Lake Formation como consumidor.

A continuación se muestra la sintaxis de la cadena del parámetro IAM_ROLE para un único ARN.

```
IAM_ROLE 'arn:aws:iam::<aws-account-id>:role/<role-name>'
```

Puede encadenar roles para que el clúster pueda asumir otro rol de IAM, que posiblemente pertenezca a otra cuenta. Puede encadenar hasta 10 roles. Para obtener más información, consulte Encadenamiento de roles de IAM en Amazon Redshift Spectrum.

Para este rol de IAM; asocie una política de permisos de IAM similar a la siguiente.

```
{
```

```
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AccessSecret",
            "Effect": "Allow",
            "Action": [
                "secretsmanager:GetResourcePolicy",
                "secretsmanager:GetSecretValue",
                "secretsmanager:DescribeSecret",
                "secretsmanager:ListSecretVersionIds"
            ],
            "Resource": "arn:aws:secretsmanager:us-west-2:123456789012:secret:my-
rds-secret-VNenFy"
        },
        {
            "Sid": "VisualEditor1",
            "Effect": "Allow",
            "Action": [
                "secretsmanager:GetRandomPassword",
                "secretsmanager:ListSecrets"
            ],
            "Resource": "*"
        }
    ]
}
```

Para obtener información sobre los pasos para crear un rol de IAM que se utilizará con la consulta federada, consulte Creación de un secreto y rol de IAM para utilizar consultas federadas.

> **Note**
>
> No incluya espacios en la lista de roles encadenados.

A continuación se muestra la sintaxis para encadenar tres roles.

```
IAM_ROLE 'arn:aws:iam::<aws-account-id>:role/<role-1-name>,arn:aws:iam::<aws-
account-id>:role/<role-2-name>,arn:aws:iam::<aws-account-id>:role/<role-3-name>'
```

## Sintaxis para utilizar CREATE DATABASE con un datashare

La siguiente sintaxis describe el comando CREATE DATABASE que se utiliza para crear bases de datos a partir de un recurso compartido de datos para compartir datos dentro de la misma cuenta de AWS.

```
CREATE DATABASE database_name
[ [ WITH PERMISSIONS ] FROM DATASHARE datashare_name OF [ ACCOUNT account_id ]
 NAMESPACE namespace_guid
```

La siguiente sintaxis describe el comando CREATE DATABASE que se utiliza para crear bases de datos a partir de un recurso compartido de datos, para compartir datos entre cuentas de AWS.

```
CREATE DATABASE database_name
[ [ WITH PERMISSIONS ] FROM DATASHARE datashare_name OF ACCOUNT account_id
 NAMESPACE namespace_guid
```

Parámetros para utilizar CREATE DATABASE con un datashare

FROM DATASHARE

Se trata de una palabra clave que indica dónde está ubicado el datashare.

datashare_name

Se trata del nombre del datashare en el que se crea la base de datos consumidora.

WITH PERMISSIONS

Especifica que la base de datos creada a partir del recurso compartido de datos requiere permisos de nivel de objeto para acceder a los objetos individuales de la base de datos. Sin esta cláusula, los usuarios o roles a los que se conceda el permiso USAGE en la base de datos tendrán acceso automáticamente a todos los objetos de la base de datos.

NAMESPACE namespace_guid

Se trata de un valor que especifica el espacio de nombres productor al que pertenece el recurso compartido de datos.

ACCOUNT account_id

Se trata de un valor que especifica la cuenta productora a la que pertenece el recurso compartido de datos.

## Notas de uso para CREATE DATABASE en el uso compartido de datos

Como superusuario de base de datos, cuando utilice CREATE DATABASE para crear bases de datos a partir de recursos compartidos de datos en la cuenta de AWS, especifique la opción NAMESPACE. La opción ACCOUNT es opcional. Cuando utilice CREATE DATABASE para crear bases de datos a partir de recursos compartidos de datos entre las cuentas de AWS, especifique las opciones ACCOUNT y NAMESPACE desde el productor.

Solo puede crear una base de datos consumidora para un recurso compartido de datos en un clúster consumidor. No se pueden crear varias bases de datos consumidoras que hagan referencia al mismo recurso compartido de datos.

## CREATE DATABASE desde AWS Glue Data Catalog

Para crear una base de datos mediante un ARN de base de datos de AWS Glue, especifique el ARN en el comando CREATE DATABASE.

```
CREATE DATABASE sampledb FROM ARN <glue-database-arn> WITH NO DATA CATALOG SCHEMA;
```

Si lo desea, también puede suministrar un valor en el parámetro IAM_ROLE. Para obtener más información sobre el parámetro y los valores aceptados, consulte Parámetros.

A continuación, se muestran ejemplos que demuestran cómo crear una base de datos a partir de un ARN mediante un rol de IAM.

```
CREATE DATABASE sampledb FROM ARN <glue-database-arn> WITH NO DATA CATALOG SCHEMA
 IAM_ROLE <iam-role-arn>
```

```
CREATE DATABASE sampledb FROM ARN <glue-database-arn> WITH NO DATA CATALOG SCHEMA
 IAM_ROLE default;
```

También puede crear una base de datos mediante DATA CATALOG SCHEMA.

```
CREATE DATABASE sampledb FROM ARN <glue-database-arn> WITH DATA CATALOG SCHEMA
 <sample_schema> IAM_ROLE default;
```

## Creación de bases de datos para recibir los resultados de las integraciones sin ETL

Para crear una base de datos mediante una identidad de integración sin ETL, especifique `integration_id` en el comando CREATE DATABASE.

```
CREATE DATABASE destination_db_name FROM INTEGRATION 'integration_id';
```

Por ejemplo, primero, recupere los identificadores de integración de SVV_INTEGRATION;

```
SELECT integration_id FROM SVV_INTEGRATION;
```

A continuación, utilice uno de los identificadores de integración recuperados para crear la base de datos que recibe integraciones sin ETL.

```
CREATE DATABASE sampledb FROM INTEGRATION 'a1b2c3d4-5678-90ab-cdef-EXAMPLE11111';
```

Si se necesita la base de datos de origen de integraciones sin ETL, por ejemplo, especifique.

```
CREATE DATABASE sampledb FROM INTEGRATION 'a1b2c3d4-5678-90ab-cdef-EXAMPLE11111'
 DATABASE sourcedb;
```

También puede establecer un intervalo de actualización para la base de datos. Por ejemplo, para establecer el intervalo de actualización en 7200 segundos para los datos de un origen de integración sin ETL:

```
CREATE DATABASE myacct_mysql FROM INTEGRATION 'a1b2c3d4-5678-90ab-cdef-EXAMPLE11111'
 SET REFRESH_INTERVAL 7200;
```

Consulte la vista de catálogo de SVV_INTEGRATION para obtener información sobre una integración sin ETL, como integration_id, target_database, origen, refresh_interval, etc.

```
SELECT * FROM svv_integration;
```

En el siguiente ejemplo se crea una base de datos a partir de una integración con el modo de historial activado.

```
CREATE DATABASE sample_integration_db FROM INTEGRATION 'a1b2c3d4-5678-90ab-cdef-
EXAMPLE11111' SET HISTORY_MODE = true;
```

## Límites de CREATE DATABASE

Amazon Redshift aplica estos límites para las bases de datos:

- Máximo de 60 bases de datos definidas por el usuario por clúster.

- Máximo de 127 bytes para un nombre de base de datos.
- El nombre de una base de datos no puede ser una palabra reservada.

## Intercalación de bases de datos

La intercalación representa un conjunto de reglas que define la forma en que el motor de base de datos compara y ordena los datos del tipo caracteres en SQL. La intercalación sin distinción entre mayúsculas y minúsculas es la utilizada con más frecuencia. Amazon Redshift utiliza intercalación sin distinción entre mayúsculas y minúsculas para facilitar la migración desde otros sistemas de almacenamiento de datos. Gracias a la compatibilidad nativa con la intercalación sin distinción entre mayúsculas y minúsculas, Amazon Redshift sigue utilizando importantes métodos de ajuste u optimización, como claves de distribución, claves de ordenación o análisis de rango restringido.

La cláusula COLLATE especifica la intercalación predeterminada para todas las columnas CHAR y VARCHAR de la base de datos. Si se especifica CASE_INSENSITIVE, todas las columnas CHAR o VARCHAR utilizan la intercalación sin distinción entre mayúsculas y minúsculas. Para obtener información acerca de la intercalación, consulte Secuencias de intercalación.

Los datos insertados o capturados en columnas donde no se distingue entre mayúsculas y minúsculas conservarán el formato original. No obstante, todas las operaciones de cadenas basadas en la comparación, incluidas la ordenación y la agrupación, no distinguirán entre mayúsculas y minúsculas. Las operaciones de búsqueda de coincidencias de patrones, como los predicados LIKE, similares a las funciones de expresión comunes y las propias funciones tampoco distinguirán entre mayúsculas y minúsculas.

Las siguientes operaciones de SQL son compatibles con la semántica de intercalación aplicable:

- operadores de comparación: **=**, **<>**, **<**, **<=**, **>**, **>=**
- operador LIKE
- cláusulas ORDER BY
- cláusulas GROUP BY
- funciones de agrupación que utilizan la comparación de cadenas, como MIN, MAX y LISTAGG
- funciones de ventana, como las cláusulas PARTITION BY y las cláusulas ORDER BY
- funciones escalares greatest() y least(), STRPOS(), REGEXP_COUNT(), REGEXP_REPLACE(), REGEXP_INSTR(), REGEXP_SUBSTR()
- cláusula Distinct

- UNION, INTERSECT y EXCEPT
- IN LIST

Para las consultas externas, incluidas las consultas federadas de Amazon Redshift Spectrum y Aurora PostgreSQL, la intercalación de la columna VARCHAR o CHAR es la misma que la intercalación actual en el nivel de base de datos.

En el siguiente ejemplo, se consulta una tabla de Amazon Redshift Spectrum:

```
SELECT ci_varchar FROM spectrum.test_collation
WHERE ci_varchar = 'AMAZON';

ci_varchar
----------
amazon
Amazon
AMAZON
AmaZon
(4 rows)
```

Para obtener información sobre cómo crear tablas mediante la intercalación de bases de datos, consulte CREATE TABLE.

Para obtener información acerca de la función COLLATE, consulte Función COLLATE.

Limitaciones de la intercalación de bases de datos

A continuación, se describen las limitaciones que existen a la hora de trabajar con la intercalación de bases de datos en Amazon Redshift:

- Todas las tablas o las vistas de sistema, incluidas las tablas del catálogo de PG y las tablas de sistema de Amazon Redshift, distinguen entre mayúsculas y minúsculas.
- Cuando la base de datos consumidora y la base de datos productora tienen intercalaciones diferentes en el nivel de base de datos, Amazon Redshift no admite consultas entre bases de datos y clústeres.
- Amazon Redshift no admite la intercalación sin distinción entre mayúsculas y minúsculas en la consulta del nodo principal específicamente.

  En el siguiente ejemplo, se muestra una consulta no admitida que no distingue entre mayúsculas y minúsculas, además del error que envía Amazon Redshift:

```
SELECT collate(usename, 'case_insensitive') FROM pg_user;
ERROR:  Case insensitive collation is not supported in leader node only query.
```

- Amazon Redshift no admite la interacción entre columnas donde se distingue entre mayúsculas y minúsculas, y columnas donde no se hace esa distinción, como comparación, función, combinación u operaciones con conjuntos.

  Los siguientes ejemplos muestran errores que ocurren cuando interactúan las columnas donde se distingue entre mayúsculas y minúsculas, y las columnas donde no se hace esa distinción:

```
CREATE TABLE test
  (ci_col varchar(10) COLLATE case_insensitive,
   cs_col varchar(10) COLLATE case_sensitive,
   cint int,
   cbigint bigint);
```

```
SELECT ci_col = cs_col FROM test;
ERROR:  Query with different collations is not supported yet.
```

```
SELECT concat(ci_col, cs_col) FROM test;
ERROR:  Query with different collations is not supported yet.
```

```
SELECT ci_col FROM test UNION SELECT cs_col FROM test;
ERROR:  Query with different collations is not supported yet.
```

```
SELECT * FROM test a, test b WHERE a.ci_col = b.cs_col;
ERROR:  Query with different collations is not supported yet.
```

```
Select Coalesce(ci_col, cs_col) from test;
ERROR:  Query with different collations is not supported yet.
```

```
Select case when cint > 0 then ci_col else cs_col end from test;
ERROR:  Query with different collations is not supported yet.
```

- Amazon Redshift no admite la intercalación para el tipo de datos SUPER. No se admiten la creación de columnas SUPER en bases de datos donde no se distingue entre mayúsculas y

minúsculas, ni las interacciones entre columnas SUPER y columnas donde no se distingue entre mayúsculas y minúsculas.

En el siguiente ejemplo, se crea una tabla con SUPER como tipo de datos en la base de datos donde no se distingue entre mayúsculas y minúsculas:

```
CREATE TABLE super_table (a super);
ERROR: SUPER column is not supported in case insensitive database.
```

En el siguiente ejemplo, se consultan datos con una cadena donde no se distingue entre mayúsculas y minúsculas, con una comparación con los datos SUPER:

```
CREATE TABLE test_super_collation
(s super, c varchar(10) COLLATE case_insensitive, i int);
```

```
SELECT s = c FROM test_super_collation;
ERROR:  Coercing from case insensitive string to SUPER is not supported.
```

Para que estas consultas funcionen, utilice la función COLLATE para convertir la intercalación de una columna de manera que coincida con la otra. Para obtener más información, consulte Función COLLATE.

## Ejemplos

### Creación de una base de datos

En el siguiente ejemplo, se crea una base de datos llamada TICKIT y se concede la propiedad al usuario DWUSER.

```
create database tickit
with owner dwuser;
```

Consulte la tabla de catálogo PG_DATABASE_INFO para ver los detalles de las bases de datos.

```
select datname, datdba, datconnlimit
from pg_database_info
where datdba > 1;

 datname     | datdba | datconnlimit
```

```
-------------+--------+-------------
 admin       |    100 | UNLIMITED
 reports     |    100 | 100
 tickit      |    100 | 100
```

En el siguiente ejemplo se crea una base de datos denominada **sampledb** con nivel de aislamiento SNAPSHOT.

```
CREATE DATABASE sampledb ISOLATION LEVEL SNAPSHOT;
```

En el siguiente ejemplo, se crea la base de datos sales_db a partir del recurso compartido de datos SalesShare.

```
CREATE DATABASE sales_db FROM DATASHARE salesshare OF NAMESPACE
 '13b8833d-17c6-4f16-8fe4-1a018f5ed00d';
```

Ejemplos de intercalación de bases de datos

Creación de una base de datos donde no se distingue entre mayúsculas y minúsculas

En el siguiente ejemplo, se crea la base de datos `sampledb`, se crea la tabla `T1` y se insertan datos en la tabla `T1`.

```
create database sampledb collate case_insensitive;
```

Conéctese a la nueva base de datos que acaba de crear mediante su cliente SQL. Cuando utilice el editor de consultas de Amazon Redshift v2, seleccione `sampledb` en el Editor. Cuando utiliza use RSQL, puede utilizar un comando como el siguiente.

```
\connect sampledb;
```

```
CREATE TABLE T1 (
  col1 Varchar(20) distkey sortkey
);
```

```
INSERT INTO T1 VALUES ('bob'), ('john'), ('Mary'), ('JOHN'), ('Bob');
```

Luego, la consulta encuentra resultados con `John`.

```
SELECT * FROM T1 WHERE col1 = 'John';

 col1
 ------
 john
 JOHN
(2 row)
```

Ordenación sin distinción entre mayúsculas y minúsculas

En el siguiente ejemplo, se muestra la ordenación sin distinción entre mayúsculas y minúsculas con la tabla T1. La ordenación de Bob y bob, o de John y john, no es determinista, porque son iguales en la columna donde no se distingue entre mayúsculas y minúsculas.

```
SELECT * FROM T1 ORDER BY 1;

 col1
 ------
 bob
 Bob
 JOHN
 john
 Mary
(5 rows)
```

Del mismo modo, el siguiente ejemplo muestra cómo se ordena sin distinción entre mayúsculas y minúsculas con la cláusula GROUP BY. Bob y bob son iguales y pertenecen al mismo grupo. No es determinístico cuál aparece en el resultado.

```
SELECT col1, count(*) FROM T1 GROUP BY 1;

 col1 | count
 -----+------
 Mary |  1
 bob  |  2
 JOHN |  2
(3 rows)
```

Consulta con una función de ventana en columnas donde no se distingue entre mayúsculas y minúsculas

En el siguiente ejemplo, se consulta una función de ventana en una columna donde no se distingue entre mayúsculas y minúsculas.

```
SELECT col1, rank() over (ORDER BY col1) FROM T1;

 col1 | rank
 -----+------
 bob  |   1
 Bob  |   1
 john |   3
 JOHN |   3
 Mary |   5
(5 rows)
```

Consulta con la palabra clave DISTINCT

En el siguiente ejemplo, se consulta la tabla T1 con la palabra clave DISTINCT.

```
SELECT DISTINCT col1 FROM T1;

 col1
 ------
 bob
 Mary
 john
(3 rows)
```

Consulta con la cláusula UNION

En el siguiente ejemplo, se muestran los resultados de la cláusula UNION de las tablas T1 y T2.

```
CREATE TABLE T2 AS SELECT * FROM T1;
```

```
SELECT col1 FROM T1 UNION SELECT col1 FROM T2;

 col1
 ------
 john
 bob
 Mary
(3 rows)
```

# CREATE DATASHARE

Cree un datashare nuevo en la base de datos actual. El propietario de este datashare es el emisor del comando CREATE DATASHARE.

Amazon Redshift asocia cada datashare a una sola base de datos de Amazon Redshift. Solo puede agregar objetos de la base de datos asociada a un datashare. Puede crear varios datashares en la misma base de datos de Amazon Redshift.

Para obtener más información sobre los recursos compartidos de datos, consulte Uso compartido de datos en Amazon Redshift.

Para ver información sobre los datashares, utilice SHOW DATASHARES.

## Privilegios necesarios

Los siguientes privilegios son necesarios para CREATE DATASHARE:

- Superusuario
- Usuarios con el privilegio CREATE DATASHARE
- Propietario de la base de datos

## Sintaxis

```
CREATE DATASHARE datashare_name
[[SET] PUBLICACCESSIBLE [=] TRUE | FALSE ];
```

## Parámetros

datashare_name

El nombre del datashare. El nombre del datashare debe ser único en el espacio de nombres del clúster.

[[SET] PUBLICACCESSIBLE]

Se trata de una cláusula que especifica si el datashare se puede compartir con clústeres accesibles de manera pública.

El valor predeterminado de `SET PUBLICACCESSIBLE` es `FALSE`.

### Notas de uso

De manera predeterminada, el propietario del recurso compartido de datos solo posee el recurso compartido pero no los objetos dentro de este.

Solo los superusuarios y el propietario de la base de datos pueden utilizar CREATE DATASHARE y delegar privilegios ALTER a otros usuarios o grupos.

### Ejemplos

En el siguiente ejemplo, se crea el recurso compartido de datos de `salesshare`.

```
CREATE DATASHARE salesshare;
```

En el siguiente ejemplo, se crea el recurso compartido de datos de `demoshare` que administra AWS Data Exchange.

```
CREATE DATASHARE demoshare SET PUBLICACCESSIBLE TRUE, MANAGEDBY ADX;
```

## CREATE EXTERNAL FUNCTION

Crea una función definida por el usuario (UDF) escalar basada en AWS Lambda para Amazon Redshift. Para obtener más información acerca de las funciones de Lambda definidas por el usuario, consulte UDF de Lambda escalares.

### Privilegios necesarios

Los siguientes privilegios son necesarios para CREATE EXTERNAL FUNCTION:

- Superusuario
- Usuarios con el privilegio CREATE [ OR REPLACE ] EXTERNAL FUNCTION

### Sintaxis

```
CREATE [ OR REPLACE ] EXTERNAL FUNCTION external_fn_name ( [data_type] [, ...] )
RETURNS data_type
{ VOLATILE | STABLE }
LAMBDA 'lambda_fn_name'
IAM_ROLE { default | ‘arn:aws:iam::<Cuenta de AWS-id>:role/<role-name>’
RETRY_TIMEOUT milliseconds
```

```
MAX_BATCH_ROWS count
MAX_BATCH_SIZE size [ KB | MB ];
```

## Parámetros

OR REPLACE

Se trata de una cláusula que especifica que, si ya existe una función con el mismo nombre y los mismos tipos de datos de argumento de entrada o firma que este, debe reemplazarse la función existente. Solo puede reemplazar una función con una nueva función que defina un conjunto idéntico de tipos de datos. Debe ser un superusuario para reemplazar una función.

Si define una función con el mismo nombre que una función existente, pero con una firma diferente, crea una nueva función. En otras palabras, se sobrecarga el nombre de la función. Para obtener más información, consulte Sobrecarga de los nombres de función.

external_fn_name

Se trata del nombre de la función externa. Si especifica un nombre de esquema (como myschema.myfunction), la función se crea con el esquema especificado. De lo contrario, la función se crea en el esquema actual. Para obtener más información acerca de los nombres válidos, consulte Nombres e identificadores.

Recomendamos que utilice el prefijo `f_` en los nombres de todas las UDF. Amazon Redshift reserva el prefijo `f_` para los nombres de las UDF. Con el uso del prefijo `f_`, se asegura de que los nombres de las UDF no entren en conflicto con ninguno de los nombres de las funciones SQL integradas para Amazon Redshift ahora o en el futuro. Para obtener más información, consulte Prevención de conflictos al dar nombre a las UDF.

data_type

Se trata del tipo de datos para los argumentos de entrada. Para obtener más información, consulte UDF de Python escalares y UDF de Lambda escalares.

RETURNS data_type

El tipo de datos del valor que la función devuelve. El tipo de datos RETURNS puede ser cualquier tipo de datos estándar de Amazon Redshift. Para obtener más información, consulte UDF de Python escalares y UDF de Lambda escalares.

VOLATILE | STABLE

Informa al optimizador de consultas acerca de la volatilidad de la función.

Para obtener la mejor optimización, etiquete la función con la categoría de volatilidad válida más estricta. Las categorías de volatilidad, de la menos estricta a la más estricta, son las siguientes:

- VOLATILE
- STABLE

VOLATILE

Dados los mismos argumentos, la función puede devolver resultados diferentes en ejecuciones consecutivas, incluso para las filas de una única instrucción. El optimizador de consultas no puede hacer suposiciones sobre el comportamiento de una función volátil. Una consulta que utiliza una función volátil debe reevaluar la función para cada entrada.

STABLE

Dados los mismos argumentos, se garantiza que la función devuelve los mismos resultados en sucesivas llamadas procesadas en una misma instrucción. La función puede devolver resultados diferentes cuando se evoca en diferentes instrucciones. Con esta categoría, el optimizador puede reducir el número de veces que se llama a la función en una misma instrucción.

Tenga en cuenta que si el rigor elegido no es válido para la función, existe el riesgo de que el optimizador omita algunas llamadas basándose en este rigor. Esto puede provocar un conjunto de resultados incorrecto.

La cláusula IMMUTABLE no es compatible actualmente con las UDF de Lambda.

LAMBDA 'lambda_fn_name'

Se trata del nombre de la función a la que llama Amazon Redshift.

Si desea conocer los pasos necesarios para crear una función de AWS Lambda, consulte Creación de una función de Lambda con la consola en la Guía para desarrolladores de AWS Lambda.

Para obtener información acerca de los permisos necesarios para la función de Lambda, consulte Permisos de AWS Lambda en la Guía para desarrolladores de AWS Lambda.

IAM_ROLE { default | ‘arn:aws:iam::*<Cuenta de AWS-id>*:role/*<role-name>*’

Utilice la palabra clave predeterminada para que Amazon Redshift utilice el rol de IAM configurado como predeterminado y asociado al clúster cuando se ejecuta el comando CREATE EXTERNAL FUNCTION.

Utilice el nombre de recurso de Amazon (ARN), de un rol de IAM que el clúster utiliza para la autenticación y la autorización. El comando CREATE EXTERNAL FUNCTION tiene autorización para invocar funciones de Lambda a través de este rol de IAM. Si su clúster tiene un rol de IAM existente con permisos para invocar funciones de Lambda adjuntas, puede sustituir el ARN de su rol. Para obtener más información, consulte Configuración del parámetro de autorización para las UDF de Lambda.

A continuación, se muestra la sintaxis del parámetro IAM_ROLE.

```
IAM_ROLE 'arn:aws:iam::aws-account-id:role/role-name'
```

RETRY_TIMEOUT milliseconds

Se trata de la cantidad de tiempo total en milisegundos que Amazon Redshift utiliza para los retrasos en el reintento de retroceso.

En lugar de volver a intentar ejecutar inmediatamente cualquier consulta que presente error, Amazon Redshift realiza retrocesos y espera cierta cantidad de tiempo entre reintentos. Luego, Amazon Redshift vuelve a intentar realizar la solicitud para ejecutar la consulta que presenta error de nuevo hasta que la suma de todos los retrasos sea igual o superior al valor RETRY_TIMEOUT que especificó. El valor predeterminado es 20 000 milisegundos.

Cuando se invoca una función de Lambda, Amazon Redshift vuelve a intentar ejecutar las consultas que presentan errores, como `TooManyRequestsException`, `EC2ThrottledException` y `ServiceException`.

Puede configurar el parámetro RETRY_TIMEOUT en 0 milisegundos para evitar cualquier reintento de ejecución de una UDF de Lambda.

MAX_BATCH_ROWS recuento

El número máximo de filas que Amazon Redshift envía en una sola solicitud por lotes para una única invocación de lambda.

El valor mínimo de este parámetro es 1. El valor máximo es INT_MAX o 2 147 483 647.

Este parámetro es opcional. El valor predeterminado es INT_MAX o 2 147 483 647.

MAX_BATCH_SIZE tamaño [ KB | MB ]

El tamaño máximo de la carga útil de datos que Amazon Redshift envía en una sola solicitud por lotes para una única invocación de lambda.

El valor mínimo de este parámetro es 1 KB. El valor máximo es 5 MB.

El valor predeterminado de este parámetro es 5 MB.

KB y MB son opcionales. Si no establece la unidad de medida, Amazon Redshift utilizará KB de forma predeterminada.

## Notas de uso

Tenga en cuenta lo siguiente al crear UDF de Lambda:

- El orden de las llamadas a la función de Lambda en los argumentos de entrada no es fijo ni está garantizado. Puede variar de una instancia a otra de las consultas en ejecución, en función de la configuración del clúster.
- No se garantiza que las funciones se apliquen a cada argumento de entrada una vez y solo una vez. La interacción entre Amazon Redshift y AWS Lambda podría dar lugar a llamadas repetitivas con las mismas entradas.

## Ejemplos

A continuación, se proporcionan ejemplos de uso de funciones escalares de Lambda definidas por el usuario (UDF).

Ejemplo de UDF escalar de Lambda con una función Node.js de Lambda

En el siguiente ejemplo, se crea una función externa denominada `exfunc_sum` que toma dos valores enteros como argumentos de entrada. Esta función devuelve la suma como una salida expresada con un número entero. El nombre de la función de Lambda a la que se llamará es `lambda_sum`. El lenguaje utilizado para esta función de Lambda es Node.js 12.x. Asegúrese de especificar el rol de IAM. El ejemplo utiliza `'arn:aws:iam::123456789012:user/johndoe'` como rol de IAM.

```
CREATE EXTERNAL FUNCTION exfunc_sum(INT,INT)
RETURNS INT
VOLATILE
LAMBDA 'lambda_sum'
IAM_ROLE 'arn:aws:iam::123456789012:role/Redshift-Exfunc-Test';
```

La función de Lambda toma la carga de solicitudes e itera sobre cada fila. Se suman todos los valores de una sola fila para calcular la suma de esa fila, que se guarda en la matriz de respuesta. La cantidad de filas en la matriz de resultados es similar a la cantidad de filas recibidas en la carga de la solicitud.

La carga de respuesta JSON debe tener los datos de resultado en el campo “results” para que la función externa la reconozca. El campo de argumentos de la solicitud enviada a la función de Lambda contiene la carga de datos. En el caso de una solicitud por lote, es posible que existan varias filas en la carga de datos. La siguiente función de Lambda itera sobre todas las filas de la carga de datos de la solicitud. También itera de forma individual sobre todos los valores dentro de una sola fila.

```
exports.handler = async (event) => {
    // The 'arguments' field in the request sent to the Lambda function contains the
 data payload.
    var t1 = event['arguments'];

    // 'len(t1)' represents the number of rows in the request payload.
    // The number of results in the response payload should be the same as the number
 of rows received.
    const resp = new Array(t1.length);

    // Iterating over all the rows in the request payload.
    for (const [i, x] of t1.entries())
    {
        var sum = 0;
        // Iterating over all the values in a single row.
        for (const y of x) {
            sum = sum + y;
        }
        resp[i] = sum;
    }
    // The 'results' field should contain the results of the lambda call.
    const response = {
        results: resp
    };
    return JSON.stringify(response);
};
```

En el siguiente ejemplo, se llama a la función externa con valores literales.

```
select exfunc_sum(1,2);
```

```
exfunc_sum
------------
 3
(1 row)
```

En el siguiente ejemplo, se crea una tabla denominada t_sum con dos columnas, c1 y c2, del tipo de datos entero y se insertan dos filas de datos. Luego, se llama a la función externa transmitiendo los nombres de las columnas de esta tabla. Las dos filas de la tabla se envían en una solicitud por lote en la carga de la solicitud como una sola invocación de Lambda.

```
CREATE TABLE t_sum(c1 int, c2 int);
INSERT INTO t_sum VALUES (4,5), (6,7);
SELECT exfunc_sum(c1,c2) FROM t_sum;
 exfunc_sum
---------------
 9
 13
(2 rows)
```

Ejemplo de UDF escalar de Lambda con el atributo RETRY_TIMEOUT

En la siguiente sección, encontrará un ejemplo de cómo utilizar el atributo RETRY_TIMEOUT en las UDF de Lambda.

Las funciones de AWS Lambda tienen límites de simultaneidad que se pueden establecer para cada función. Para obtener más información acerca de los límites de simultaneidad, consulte Administración de la simultaneidad para una función de Lambda en la Guía para desarrolladores de AWS Lambda y la publicación Administración de la simultaneidad de la función AWS Lambda en el blog de informática de AWS.

Cuando la cantidad de solicitudes que atiende una UDF de Lambda supera los límites de simultaneidad, las solicitudes nuevas reciben el error `TooManyRequestsException`. La UDF de Lambda intenta solucionar este error de nuevo hasta que la suma de todos los retrasos entre las solicitudes enviadas a la función de Lambda sea igual o superior al valor RETRY_TIMEOUT que estableció. El valor predeterminado de RETRY_TIMEOUT es 20 000 milisegundos.

En el siguiente ejemplo, se utiliza una función de Lambda denominada `exfunc_sleep_3`. Esta función toma la carga de la solicitud, itera sobre cada fila y convierte la entrada a mayúsculas. Luego, está inactiva durante 3 segundos y devuelve el resultado. El lenguaje utilizado para esta función de Lambda es Python 3.8.

La cantidad de filas en la matriz de resultados es similar a la cantidad de filas recibidas en la carga de la solicitud. La carga de respuesta JSON debe tener los datos de resultado en el campo `results` para que la función externa la reconozca. El campo de `arguments` de la solicitud enviada a la función de Lambda contiene la carga de datos. En el caso de una solicitud por lote, es posible que aparezcan varias filas en la carga de datos.

El límite de simultaneidad para esta función se configura específicamente en 1 en simultaneidad reservada para demostrar el uso del atributo RETRY_TIMEOUT. Cuando el atributo se configura en 1, la función de Lambda solo puede atender una solicitud a la vez.

```
import json
import time
def lambda_handler(event, context):
    t1 = event['arguments']
    # 'len(t1)' represents the number of rows in the request payload.
    # The number of results in the response payload should be the same as the number of
 rows received.
    resp = [None]*len(t1)

    # Iterating over all rows in the request payload.
    for i, x in enumerate(t1):
        # Iterating over all the values in a single row.
        for j, y in enumerate(x):
            resp[i] = y.upper()

    time.sleep(3)
    ret = dict()
    ret['results'] = resp
    ret_json = json.dumps(ret)
    return ret_json
```

A continuación, se incluyen dos ejemplos adicionales que ilustran el atributo RETRY_TIMEOUT. Cada ejemplo invoca una sola UDF de Lambda. Cuando invocan la UDF de Lambda, cada ejemplo ejecuta la misma consulta SQL para invocar la UDF de Lambda desde dos sesiones de base de datos simultáneas. Cuando la primera consulta que invoca la UDF de Lambda es atendida por la UDF, la segunda consulta recibe el error `TooManyRequestsException`. Este resultado se produce porque se configura de forma específica la simultaneidad reservada en 1 en la UDF. Para obtener información sobre cómo establecer la simultaneidad reservada para las funciones de Lambda, consulte Configuración de la simultaneidad reservada.

A continuación, el primer ejemplo configura el atributo RETRY_TIMEOUT para la UDF de Lambda en 0 milisegundos. Si la solicitud de Lambda recibe alguna excepción de la función de Lambda, Amazon Redshift no realiza ningún reintento. Este resultado se produce porque el atributo RETRY_TIMEOUT se configura en 0.

```
CREATE OR REPLACE EXTERNAL FUNCTION exfunc_upper(varchar)
RETURNS varchar
VOLATILE
LAMBDA 'exfunc_sleep_3'
IAM_ROLE 'arn:aws:iam::123456789012:role/Redshift-Exfunc-Test'
RETRY_TIMEOUT 0;
```

Con RETRY_TIMEOUT configurado en 0, puede ejecutar las siguientes dos consultas desde sesiones de base de datos independientes para ver resultados diferentes.

La primera consulta SQL que utiliza la UDF de Lambda se ejecuta correctamente.

```
select exfunc_upper('Varchar');
 exfunc_upper
 --------------
 VARCHAR
(1 row)
```

La segunda consulta, que se ejecuta desde una sesión de base de datos independiente al mismo tiempo, recibe el error `TooManyRequestsException`.

```
select exfunc_upper('Varchar');
ERROR:  Rate Exceeded.; Exception: TooManyRequestsException; ShouldRetry: 1
DETAIL:
-----------------------------------------------
error:  Rate Exceeded.; Exception: TooManyRequestsException; ShouldRetry: 1
code:      32103
context:query:     0
location:  exfunc_client.cpp:102
process:   padbmaster [pid=26384]
-----------------------------------------------
```

A continuación, el segundo ejemplo configura el atributo RETRY_TIMEOUT para la UDF de Lambda en 3000 milisegundos. Incluso si la segunda consulta se ejecuta de forma simultánea, la UDF de Lambda reintenta la ejecución hasta que el retraso total sea de 3000 milisegundos. Por lo tanto, ambas consultas se ejecutan correctamente.

```
CREATE OR REPLACE EXTERNAL FUNCTION exfunc_upper(varchar)
RETURNS varchar
VOLATILE
LAMBDA 'exfunc_sleep_3'
IAM_ROLE 'arn:aws:iam::123456789012:role/Redshift-Exfunc-Test'
RETRY_TIMEOUT 3000;
```

Con RETRY_TIMEOUT configurado en 3000, puede ejecutar las siguientes dos consultas desde sesiones de base de datos independientes para ver los mismos resultados.

La primera consulta SQL que ejecuta la UDF de Lambda se ejecuta correctamente.

```
select exfunc_upper('Varchar');
 exfunc_upper
 --------------
 VARCHAR
(1 row)
```

La segunda consulta se ejecuta de forma simultánea, y la UDF de Lambda reintenta la ejecución hasta que el retraso total sea de 3000 milisegundos.

```
select exfunc_upper('Varchar');
 exfunc_upper
--------------
 VARCHAR
(1 row)
```

Ejemplo de UDF escalar de Lambda con una función Python de Lambda

En el siguiente ejemplo, se crea una función externa denominada `exfunc_multiplication` que multiplica números y devuelve un número entero. Este ejemplo incorpora los campos de éxito y `error_msg` en la respuesta de Lambda. El campo de éxito se establece como false cuando hay un desbordamiento de valores enteros en el resultado de la multiplicación y el mensaje `error_msg` se establece como `Integer multiplication overflow`. La función `exfunc_multiplication` toma tres valores enteros como argumentos de entrada y devuelve la suma como un valor de salida entero.

El nombre de la función de Lambda que se llama es `lambda_multiplication`. El lenguaje utilizado para esta función de Lambda es Python 3.8. Asegúrese de especificar el rol de IAM.

```
CREATE EXTERNAL FUNCTION exfunc_multiplication(int, int, int)
```

```
RETURNS INT
VOLATILE
LAMBDA 'lambda_multiplication'
IAM_ROLE 'arn:aws:iam::123456789012:role/Redshift-Exfunc-Test';
```

La función de Lambda toma la carga de solicitudes e itera sobre cada fila. Todos los valores de una sola fila se multiplican para calcular el resultado de esa fila, que se guarda en la lista de respuestas. En este ejemplo, se utiliza un valor booleano de éxito establecido como true de manera predeterminada. Si el resultado de la multiplicación de una fila tiene un desbordamiento de valores enteros, el valor de éxito se establece como false. Por lo tanto, se interrumpe el ciclo de iteraciones.

Durante la creación de la carga de respuesta, si el valor de éxito es false, la siguiente función de Lambda agrega el campo `error_msg` a la carga. También establece el mensaje de error en `Integer multiplication overflow`. Si el valor de éxito es true, los datos de resultado se agregan en el campo de resultados. La cantidad de filas en la matriz de resultados, si hay, es similar a la cantidad de filas recibidas en la carga de la solicitud.

El campo de argumentos de la solicitud enviada a la función de Lambda contiene la carga de datos. En el caso de una solicitud por lote, es posible que existan varias filas en la carga de datos. La siguiente función de Lambda itera sobre todas las filas de la carga de datos de la solicitud e itera de manera individual sobre todos los valores dentro de una sola fila.

```
import json
def lambda_handler(event, context):
    t1 = event['arguments']
    # 'len(t1)' represents the number of rows in the request payload.
    # The number of results in the response payload should be the same as the number of
 rows received.
    resp = [None]*len(t1)

    # By default success is set to 'True'.
    success = True
    # Iterating over all rows in the request payload.
    for i, x in enumerate(t1):
        mul = 1
        # Iterating over all the values in a single row.
        for j, y in enumerate(x):
            mul = mul*y

        # Check integer overflow.
        if (mul >= 9223372036854775807 or mul <= -9223372036854775808):
```

```
            success = False
            break
        else:
            resp[i] = mul
    ret = dict()
    ret['success'] = success
    if not success:
        ret['error_msg'] = "Integer multiplication overflow"
    else:
        ret['results'] = resp
    ret_json = json.dumps(ret)

    return ret_json
```

En el siguiente ejemplo, se llama a la función externa con valores literales.

```
SELECT exfunc_multiplication(8, 9, 2);
  exfunc_multiplication
---------------------------
          144
(1 row)
```

En el siguiente ejemplo, se crea una tabla denominada t_multi con tres columnas, c1, c2 y c3, del tipo de datos entero. Se llama a la función externa transmitiendo los nombres de las columnas de esta tabla. Los datos se insertan de tal manera que provocan un desbordamiento de valores enteros para mostrar cómo se propaga el error.

```
CREATE TABLE t_multi (c1 int, c2 int, c3 int);
INSERT INTO t_multi VALUES (2147483647, 2147483647, 4);
SELECT exfunc_multiplication(c1, c2, c3) FROM t_multi;
DETAIL:
  -----------------------------------------------
  error:  Integer multiplication overflow
  code:      32004context:
  context:
  query:     38
  location:  exfunc_data.cpp:276
  process:   query2_16_38 [pid=30494]
  -----------------------------------------------
```

# CREATE EXTERNAL MODEL

Temas


- Requisitos previos para CREATE EXTERNAL MODEL
- Privilegios necesarios
- Control de costes
- Sintaxis de CREATE EXTERNAL MODEL
- Parámetros y configuración de CREATE EXTERNAL MODEL
- Parámetros de función de inferencia de CREATE EXTERNAL MODEL


## Requisitos previos para CREATE EXTERNAL MODEL

Antes de utilizar la instrucción CREATE EXTERNAL MODEL, complete los requisitos previos en Configuración del clúster para utilizar Amazon Redshift ML. A continuación, se brinda un resumen de alto nivel de los requisitos previos.

- Cree un clúster de Amazon Redshift con la consola de administración de AWS o la AWS Command Line Interface (AWS CLI)
- Adjunte la política de AWS Identity and Access Management (IAM) mientras crea el clúster.
- Para permitir que Amazon Redshift y Amazon Bedrock asuman el rol a la hora de interactuar con otros servicios, agregue la política de confianza adecuada al rol de IAM.
- Habilite el acceso a los LLM específicos que desee usar desde la consola de Amazon Bedrock.
- (Opcional) Si encuentra excepciones que imponen limitaciones provenientes de Amazon Bedrock, por ejemplo, `Too many requests, please wait before trying again`, incluso con pequeñas cantidades de datos, compruebe las cuotas en Service Quotas en su cuenta de Amazon Bedrock. Compruebe que la cuota de cuenta aplicada sea como mínimo igual al valor de la cuota predeterminada de AWS para las solicitudes de InvokeModel para el modelo que está utilizando.

Para obtener detalles sobre el rol de IAM, la política de confianza y otros requisitos previos, consulte Configuración del clúster para utilizar Amazon Redshift ML.

## Privilegios necesarios

Los siguientes privilegios son necesarios para CREATE EXTERNAL MODEL:

- Superusuario
- Usuarios con el privilegio CREATE MODEL
- Roles con el privilegio GRANT CREATE MODEL

## Control de costes

Amazon Redshift ML utiliza los recursos de clúster existentes para crear modelos de predicción, por lo que no es necesario pagar costes adicionales. Sin embargo, los cargos de AWS por el uso de Amazon Bedrock dependen del modelo que seleccione. Para obtener más información, consulte Costes para usar Amazon Redshift ML.

## Sintaxis de CREATE EXTERNAL MODEL

A continuación, se muestra la sintaxis completa de la instrucción CREATE EXTERNAL MODEL.

```
CREATE EXTERNAL MODEL model_name
FUNCTION function_name
IAM_ROLE {default/'arn:aws:iam::<account-id>:role/<role-name>'}
MODEL_TYPE BEDROCK
SETTINGS (
   MODEL_ID model_id
   [, PROMPT 'prompt prefix']
   [, SUFFIX 'prompt suffix']
   [, REQUEST_TYPE {RAW|UNIFIED}]
   [, RESPONSE_TYPE {VARCHAR|SUPER}]
);
```

El comando `CREATE EXTERNAL MODEL` crea una función de inferencia que se utiliza para generar contenido.

A continuación, se muestra la sintaxis de una función de inferencia que `CREATE EXTERNAL MODEL` crea mediante un `REQUEST_TYPE` de `RAW`:

```
SELECT inference_function_name(request_super)
[FROM table];
```

A continuación, se muestra la sintaxis de una función de inferencia que `CREATE EXTERNAL MODEL` crea mediante un `REQUEST_TYPE` de `UNIFIED`:

```
SELECT inference_function_name(input_text, [, inference_config [,
 additional_model_request_fields]])
[FROM table];
```

Para obtener información sobre cómo usar la función de inferencia, consulte Uso de un modelo externo para la integración de Amazon Redshift ML con Amazon Bedrock.

## Parámetros y configuración de CREATE EXTERNAL MODEL

En esta sección se describen los parámetros y la configuración del comando `CREATE EXTERNAL MODEL`.

Temas

- Parámetros de CREATE EXTERNAL MODEL
- Configuración de CREATE EXTERNAL MODEL

### Parámetros de CREATE EXTERNAL MODEL

model_name

El nombre del modelo externo. El nombre del modelo en un esquema debe ser único.

FUNCTION function_name (data_type [,...] )

El nombre de la función de inferencia que `CREATE EXTERNAL MODEL` crea. La función de inferencia se utiliza para enviar solicitudes a Amazon Bedrock y recuperar el texto generado por ML.

IAM_ROLE { default | 'arn:aws:iam::<account-id>:role/<role-name>' }

El rol de IAM que Amazon Redshift utiliza para acceder a Amazon Bedrock. Para obtener información acerca del rol de IAM, consulte Creación o actualización de un rol de IAM para la integración de Amazon Redshift ML con Amazon Bedrock.

MODEL_TYPE BEDROCK

Especifica el tipo de modelo. El único valor válido es `BEDROCK`.

SETTINGS ( MODEL_ID model_id [,...] )

Especifica la configuración del modelo externo. Consulte la siguiente sección para obtener más información.

Configuración de CREATE EXTERNAL MODEL

MODEL_ID model_id

El identificador del modelo externo, por ejemplo, `anthropic.claude-v2`. Para obtener información sobre los ID de modelo de Amazon Bedrock, consulte Amazon Bedrock model IDs.

PROMPT “prefijo de petición”

Especifica una petición estática que Amazon Redshift agrega al principio de cada solicitud de inferencia. Solo se admite con un `REQUEST_TYPE` de `UNIFIED`.

SUFFIX “sufijo de petición”

Especifica una petición estática que Amazon Redshift agrega al final de cada solicitud de inferencia. Solo se admite con un `REQUEST_TYPE` de `UNIFIED`.

REQUEST_TYPE { RAW | UNIFIED }

Especifica el formato de la solicitud enviada a Amazon Bedrock. Entre los valores válidos se incluyen los siguientes:

- RAW: la función de inferencia toma la entrada como un supervalor único y siempre devuelve un supervalor. El formato del supervalor es específico del modelo de Amazon Bedrock seleccionado. Un súper es un modelo de predicción que combina varios algoritmos para producir una única predicción mejorada.
- UNIFIED: la función de inferencia utiliza la API unificada. Todos los modelos tienen una interfaz unificada y coherente con Amazon Bedrock. Esto funciona para todos los modelos que admiten mensajes. Este valor es el valor predeterminado.

  Para obtener más información, consulte la Documentación de la API de conversión en la Documentación de la API de Amazon Bedrock.

RESPONSE_TYPE { VARCHAR | SUPER }

Especifica el formato de la respuesta. Si `REQUEST_TYPE` es `RAW`, `RESPONSE_TYPE` es obligatorio y el único valor válido es `SUPER`. Para todos los demás valores `REQUEST  TYPE`, el valor predeterminado es `VARCHAR` y `RESPONSE_TYPE` es opcional. Entre los valores válidos se incluyen los siguientes:

- VARCHAR: Amazon Redshift solo devuelve la respuesta de texto generada por el modelo.
- SUPER: Amazon Redshift devuelve todo el JSON de respuesta generado por el modelo en forma de súper. Esto incluye la respuesta de texto e información como el motivo de la parada y

el uso de los tokens de entrada y salida del modelo. Un súper es un modelo de predicción que combina varios algoritmos para producir una única predicción mejorada.

### Parámetros de función de inferencia de CREATE EXTERNAL MODEL

En esta sección se describen los parámetros válidos para la función de inferencia que crea el comando `CREATE EXTERNAL MODEL`.

Parámetros de función de inferencia de CREATE EXTERNAL MODEL para **`REQUEST_TYPE`** de **`RAW`**

Una función de inferencia creada con un `REQUEST_TYPE` de `RAW` tiene un argumento de superentrada y siempre devuelve un tipo de superdatos. La sintaxis de la superentrada sigue la sintaxis de la solicitud del modelo específico seleccionado de Amazon Bedrock.

Parámetros de función de inferencia de CREATE EXTERNAL MODEL para **`REQUEST_TYPE`** de **`UNIFIED`**

input_text

El texto que Amazon Redshift envía a Amazon Bedrock.

inference_config

Un supervalor que contiene parámetros opcionales que Amazon Redshift envía a Amazon Bedrock. Estos pueden incluir los siguientes:

- maxTokens
- stopSequences
- temperature
- topP

Todos estos parámetros son opcionales y distinguen entre mayúsculas y minúsculas. Para obtener información sobre estos parámetros, consulte InferenceConfiguration en la referencia de la API de Amazon Bedrock.

## CREATE EXTERNAL SCHEMA

Crea un nuevo esquema externo en la base de datos actual. Puede utilizar este esquema externo para conectarse a las bases de datos de Amazon RDS for PostgreSQL o la Edición compatible con PostgreSQL de Amazon Aurora. También puede crear un esquema externo que referencie una base

de datos en un catálogo de datos externo, como AWS Glue, Athena o una base de datos en un metastore de Apache Hive, como Amazon EMR.

El propietario de este esquema es el emisor del comando CREATE EXTERNAL SCHEMA. Para transferir la propiedad de un esquema externo, use ALTER SCHEMA para cambiar el propietario. Para conceder acceso al esquema a otros usuarios o grupos utilice el comando GRANT.

No puede usar los comandos GRANT o REVOKE para los permisos en una tabla externa. En lugar de ello, conceda o revoque los permisos en el esquema externo.

> **Note**
>
> Si actualmente tiene tablas externas de Redshift Spectrum en el catálogo de datos de Amazon Athena, puede migrar el catálogo de datos de Athena a un AWS Glue Data Catalog. Para utilizar el catálogo de datos de AWS Glue con Redshift Spectrum, es posible que tenga que cambiar las políticas de AWS Identity and Access Management (IAM). Para obtener más información, consulte Actualización al catálogo de datos de AWS Glue en la Guía del usuario de Athena.

Para ver detalles de los esquemas externos, consulte la vista del sistema SVV_EXTERNAL_SCHEMAS.

## Sintaxis

La sintaxis siguiente describe el comando CREATE EXTERNAL SCHEMA utilizado para hacer referencia a datos mediante un catálogo de datos externo. Para obtener más información, consulte Amazon Redshift Spectrum.

```
CREATE EXTERNAL SCHEMA [IF NOT EXISTS] local_schema_name
FROM [ [ DATA CATALOG ] | HIVE METASTORE | POSTGRES | MYSQL | KINESIS | MSK | REDSHIFT
 | KAFKA ]
[ DATABASE 'database_name' ]
[ SCHEMA 'schema_name' ]
[ REGION 'aws-region' ]
[ IAM_ROLE [ default | 'SESSION' | 'arn:aws:iam::<Cuenta de AWS-id>:role/<role-
name>' ] ]
[ AUTHENTICATION [ none | iam | mtls] ]
[ AUTHENTICATION_ARN 'acm-certificate-arn' | SECRET_ARN 'ssm-secret- arn' ]
[ URI ['hive_metastore_uri' [ PORT port_number ] | 'hostname' [ PORT port_number ] |
 'Kafka bootstrap URL'] ]
```

```
[ CLUSTER_ARN 'arn:aws:kafka:<region>:<Cuenta de AWS-id>:cluster/msk/<cluster uuid>' ]
[ CATALOG_ROLE [ 'SESSION' | 'catalog-role-arn-string' ] ]
[ CREATE EXTERNAL DATABASE IF NOT EXISTS ]
[ CATALOG_ID 'Amazon Web Services account ID containing Glue or Lake Formation
 database' ]
```

La siguiente sintaxis describe el comando CREATE EXTERNAL SCHEMA que se utiliza para referenciar datos mediante una consulta federada a RDS POSTGRES o Aurora PostgreSQL. También se puede crear un esquema externo que haga referencia a orígenes de streaming, como Kinesis Data Streams. Para obtener más información, consulte Consulta de datos con consultas federadas en Amazon Redshift.

```
CREATE EXTERNAL SCHEMA [IF NOT EXISTS] local_schema_name
FROM POSTGRES
DATABASE 'federated_database_name' [SCHEMA 'schema_name']
URI 'hostname' [ PORT port_number ]
IAM_ROLE [ default | 'arn:aws:iam::<Cuenta de AWS-id>:role/<role-name>' ]
SECRET_ARN 'ssm-secret-arn'
```

La siguiente sintaxis describe el comando CREATE EXTERNAL SCHEMA que se utiliza para referenciar datos mediante una consulta federada a RDS MySQL o Aurora MySQL. Para obtener más información, consulte Consulta de datos con consultas federadas en Amazon Redshift.

```
CREATE EXTERNAL SCHEMA [IF NOT EXISTS] local_schema_name
FROM MYSQL
DATABASE 'federated_database_name'
URI 'hostname' [ PORT port_number ]
IAM_ROLE [ default | 'arn:aws:iam::<Cuenta de AWS-id>:role/<role-name>' ]
SECRET_ARN 'ssm-secret-arn'
```

La siguiente sintaxis describe el comando CREATE EXTERNAL SCHEMA que se utiliza para hacer referencia a datos de un flujo de Kinesis. Para obtener más información, consulte Ingesta de streaming a una vista materializada.

```
CREATE EXTERNAL SCHEMA [IF NOT EXISTS] schema_name
FROM KINESIS
IAM_ROLE [ default | 'arn:aws:iam::<Cuenta de AWS-id>:role/<role-name>' ]
```

La siguiente sintaxis describe el comando CREATE EXTERNAL SCHEMA utilizado para hacer referencia al clúster de Amazon Managed Streaming para Apache Kafka o Confluent Cloud y los

temas desde los cuales se incorpora. Para conectarse, debe proporcionar el URI del agente. Para obtener más información, consulte Ingesta de streaming a una vista materializada.

```
CREATE EXTERNAL SCHEMA [IF NOT EXISTS] schema_name
FROM KAFKA
[ IAM_ROLE [ default | 'arn:aws:iam::<Cuenta de AWS-id>:role/<role-name>' ] ]
URI 'Kafka bootstrap URI'
AUTHENTICATION [ none | iam | mtls ]
[ AUTHENTICATION_ARN 'acm-certificate-arn' | SECRET_ARN 'ssm-secret- arn' ];
```

La siguiente sintaxis describe el comando CREATE EXTERNAL SCHEMA que se utiliza para referenciar datos mediante una consulta entre distintas bases de datos.

```
CREATE EXTERNAL SCHEMA local_schema_name
FROM  REDSHIFT
DATABASE 'redshift_database_name' SCHEMA 'redshift_schema_name'
```

## Parámetros

IF NOT EXISTS

Una cláusula que indica que si el esquema especificado ya existe, el comando no debe realizar cambios y debe devolver un mensaje en el que se indique que el esquema existe, en lugar de terminar con un error. Esta cláusula es útil cuando se realiza scripting, para que el script no produzca un error si CREATE EXTERNAL SCHEMA intenta crear un esquema que ya existe.

nombre_de_esquema_local

El nombre del nuevo esquema externo. Para obtener más información acerca de los nombres válidos, consulte Nombres e identificadores.

FROM [ DATA CATALOG ] | HIVE METASTORE | POSTGRES | MYSQL | KINESIS | MSK | REDSHIFT

Una palabra clave que indica dónde está ubicada la base de datos externa.

DATA CATALOG indica que la base de datos externa se define en el catálogo de datos de Athena o en el AWS Glue Data Catalog.

Si la base de datos externa se define en un catálogo de datos externo en una región de AWS diferente, el parámetro REGION es obligatorio. DATA CATALOG es el valor predeterminado.

HIVE METASTORE indica que la base de datos externa está definida en un metaalmacén Apache Hive. Si se especifica HIVE METASTORE, se requiere el URI.

POSTGRES indica que la base de datos externa está definida en RDS PostgreSQL o Aurora PostgreSQL.

MYSQL indica que la base de datos externa está definida en RDS MySQL o Aurora MySQL.

KINESIS indica que el origen de datos es un flujo de Kinesis Data Streams.

MSK indica que el origen de datos es un clúster aprovisionado o sin servidor de Amazon MSK.

KAFKA indica que el origen de datos es un clúster de Kafka. Puede utilizar esta palabra clave para Amazon MSK y Confluent Cloud.

FROM REDSHIFT

Se trata de una palabra clave que indica que la base de datos se encuentra en Amazon Redshift.

DATABASE 'redshift_database_name' SCHEMA 'redshift_schema_name'

Se trata del nombre de la base de datos de Amazon Redshift.

El parámetro redshift_schema_name indica el esquema en Amazon Redshift. El parámetro redshift_schema_name predeterminado es `public`.

DATABASE 'federated_database_name'

Se trata de una palabra clave que indica el nombre de la base de datos externa en un motor de base de datos de PostgreSQL o MySQL compatible.

[SCHEMA 'schema_name']

El parámetro schema_name indica el esquema en un motor de base de datos de PostgreSQL compatible. El valor predeterminado de schema_name es `public`.

No se puede especificar un SCHEMA cuando se configura una consulta federada a un motor de base de datos de MySQL compatible.

REGION 'aws-region'

Si la base de datos externa se define en un catálogo de datos de Athena o en el AWS Glue Data Catalog, esta es la región de AWS en la que se encuentra la base de datos. Este parámetro es obligatorio si la base de datos se define en un catálogo de datos externo.

URI [ 'hive_metastore_uri' [ PORT port_number ] | 'hostname' [ PORT port_number ] | 'Kafka bootstrap URI' ]

Se trata del URI del nombre del alojamiento y el port_number de un motor de base de datos de PostgreSQL o MySQL compatible. El hostname es el nodo principal del conjunto de réplicas. El punto de conexión debe ser accesible (enrutable) desde el clúster de Amazon Redshift. El port_number predeterminado de PostgreSQL es 5432. El port_number predeterminado para MySQL es 3306.

> **Note**
>
> El motor de base de datos de PostgreSQL o MySQL compatible debe estar en la misma VPC que el clúster de Amazon Redshift con un grupo de seguridad que vincule Amazon Redshift y RDS url-rsPostgreSQL o Aurora PostgreSQL. Además, puede usar el enrutamiento de VPC mejorado para configurar un caso de uso entre VPC. Para obtener más información, consulte Puntos de conexión de VPC administrador por Redshift.

Especificación de un URI de metastore de hive

Si la base de datos está en un metaalmacén Hive, especifique el URI y, de manera opcional, el número de puerto para el metaalmacén. El número de puerto predeterminado es 9083.

Un URI no contiene una especificación de protocolo (“http://”). Un ejemplo de URI válido es `uri '172.10.10.10'`.

Especificación de un URI de agente para la ingesta de streaming

La inclusión del URI de bootstrap-broker proporciona la capacidad de conectarse a un clúster de Amazon MSK o Confluent Cloud y recibir datos transmitidos. Para obtener más información y ver un ejemplo, consulte Introducción a la ingesta de streaming de Amazon Managed Streaming para Apache Kafka.

IAM_ROLE [ default | 'SESSION' | 'arn:aws:iam::*<Cuenta de AWS-id>*:role/*<role-name>*' ]

Utilice la palabra clave predeterminada para que Amazon Redshift utilice el rol de IAM configurado como predeterminado y asociado al clúster cuando se ejecuta el comando CREATE EXTERNAL SCHEMA.

Use `'SESSION'` si se conecta al clúster de Amazon Redshift mediante una identidad federada y acceda a las tablas desde el esquema externo creado con este comando. Para obtener más

información, consulte Uso de una identidad federada para administrar el acceso de Amazon Redshift a los recursos locales y a las tablas externas de Amazon Redshift Spectrum, lo que explica cómo configurar la identidad federada. Tenga en cuenta que esta configuración, con `'SESSION'` en lugar del ARN, solo se puede usar si el esquema se crea utilizando `DATA CATALOG`.

Utilice el nombre de recurso de Amazon (ARN), de un rol de IAM que el clúster utiliza para la autenticación y la autorización. Como mínimo, el rol de IAM debe tener permiso para realizar una operación LIST en el bucket de Amazon S3 al que se accederá y una operación GET en los objetos de Amazon S3 que el bucket contiene.

A continuación se muestra la sintaxis de la cadena del parámetro IAM_ROLE para un único ARN.

```
IAM_ROLE 'arn:aws:iam::<aws-account-id>:role/<role-name>'
```

Puede encadenar roles para que el clúster pueda asumir otro rol de IAM, que posiblemente pertenezca a otra cuenta. Puede encadenar hasta 10 roles. Para ver un ejemplo de encadenamiento de roles, consulte Encadenamiento de roles de IAM en Amazon Redshift Spectrum.

Para este rol de IAM; asocie una política de permisos de IAM similar a la siguiente.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AccessSecret",
            "Effect": "Allow",
            "Action": [
                "secretsmanager:GetResourcePolicy",
                "secretsmanager:GetSecretValue",
                "secretsmanager:DescribeSecret",
                "secretsmanager:ListSecretVersionIds"
            ],
            "Resource": "arn:aws:secretsmanager:us-west-2:123456789012:secret:my-
rds-secret-VNenFy"
        },
        {
            "Sid": "VisualEditor1",
            "Effect": "Allow",
            "Action": [
```

```
                "secretsmanager:GetRandomPassword",
                "secretsmanager:ListSecrets"
            ],
            "Resource": "*"
        }
    ]
}
```

Para obtener información sobre los pasos para crear un rol de IAM que se utilizará con la consulta federada, consulte Creación de un secreto y rol de IAM para utilizar consultas federadas.

> **Note**
>
> No incluya espacios en la lista de roles encadenados.

A continuación se muestra la sintaxis para encadenar tres roles.

```
IAM_ROLE 'arn:aws:iam::<aws-account-id>:role/<role-1-name>,arn:aws:iam::<aws-account-id>:role/<role-2-name>,arn:aws:iam::<aws-account-id>:role/<role-3-name>'
```

SECRET_ARN 'ssm-secret-arn'

Se trata del nombre de recurso de Amazon (ARN) de un secreto de motor de base de datos de PostgreSQL o MySQL compatible creado mediante AWS Secrets Manager. Para obtener información sobre cómo crear y recuperar un ARN para un secreto, consulte Administrar secretos con AWS Secrets Manager en la Guía del usuario de AWS Secrets Manager, y Recuperar el nombre de recurso de Amazon (ARN) del secreto en Amazon Redshift.

CATALOG_ROLE [ 'SESSION' | catalog-role-arn-string]

Use `'SESSION'` para conectarse al clúster de Amazon Redshift mediante una identidad federada para la autenticación y la autorización del catálogo de datos. Para obtener más información acerca de completar los pasos para la identidad federada, consulte Uso de una identidad federada para administrar el acceso de Amazon Redshift a los recursos locales y a las tablas externas de Amazon Redshift Spectrum. Tenga en cuenta que el rol `'SESSION'` solo se puede usar si el esquema se crea en el CATÁLOGO DE DATOS.

Utilice el nombre de recurso de Amazon (ARN), para un rol de IAM que el clúster utiliza para la autenticación y la autorización del catálogo de datos.

Si no se especifica CATALOG_ROLE, Amazon Redshift utiliza el IAM_ROLE especificado. El rol del catálogo debe tener permiso para acceder al catálogo de datos en AWS Glue o Athena. Para obtener más información, consulte Políticas de IAM para Amazon Redshift Spectrum.

A continuación se muestra la sintaxis de la cadena del parámetro CATALOG_ROLE para un único ARN.

```
CATALOG_ROLE 'arn:aws:iam::<aws-account-id>:role/<catalog-role>'
```

Puede encadenar roles para que el clúster pueda asumir otro rol de IAM, que posiblemente pertenezca a otra cuenta. Puede encadenar hasta 10 roles. Para obtener más información, consulte Encadenamiento de roles de IAM en Amazon Redshift Spectrum.

> **Note**
>
> La lista de roles encadenados no debe incluir espacios.

A continuación se muestra la sintaxis para encadenar tres roles.

```
CATALOG_ROLE 'arn:aws:iam::<aws-account-id>:role/<catalog-role-1-
name>,arn:aws:iam::<aws-account-id>:role/<catalog-role-2-name>,arn:aws:iam::<aws-
account-id>:role/<catalog-role-3-name>'
```

### CREATE EXTERNAL DATABASE IF NOT EXISTS

Una cláusula que crea una base de datos externa con el nombre especificado por el argumento DATABASE, si la base de datos externa especificada no existe. Si la base de datos externa especificada existe, el comando no realiza cambios. En este caso, el comando devuelve un mensaje en el que se indica que la base de datos externa existe, en lugar de terminar con un error.

> **Note**
>
> No puede utilizar CREATE EXTERNAL DATABASE IF NOT EXISTS con HIVE METASTORE.

Para utilizar CREATE EXTERNAL DATABASE IF NOT EXISTS con un catálogo de datos habilitado para AWS Lake Formation, necesita tener el permiso CREATE_DATABASE en el catálogo de datos.

CATALOG_ID "ID de la cuenta de Amazon Web Services que contiene la base de datos de Glue o Lake Formation"

El ID de la cuenta donde se almacena la base de datos del catálogo de datos.

CATALOG_ID solo se puede especificar si planea conectarse al clúster de Amazon Redshift o a Amazon Redshift sin servidor mediante una identidad federada para la autenticación y la autorización del catálogo de datos estableciendo una de las siguientes opciones:

- De CATALOG_ROLE a 'SESSION'
- IAM_ROLE a 'SESSION' y 'CATALOG_ROLE' configurado en su valor predeterminado

Para obtener más información acerca de completar los pasos para la identidad federada, consulte Uso de una identidad federada para administrar el acceso de Amazon Redshift a los recursos locales y a las tablas externas de Amazon Redshift Spectrum.

AUTENTICACIÓN

El tipo de autenticación que se ha definido para la ingesta de streaming. La ingesta de streaming con tipos de autenticación funciona con Amazon Managed Streaming para Apache Kafka. Los tipos de AUTHENTICATION son los siguientes:

- ninguno: especifica que no se requiere autenticación. Esto corresponde al acceso no autenticado en MSK o texto sin formato con TLS en Apache Kafka.
- iam: especifica la autenticación de IAM. Al elegir esta opción, asegúrese de que el rol de IAM tenga permisos para la autenticación de IAM. Para obtener más información acerca de la definición del esquema externo, consulte Introducción a la ingesta de transmisiones desde orígenes de Apache Kafka.
- mtls: especifica que la seguridad de la capa de transporte mutua proporciona una comunicación segura al facilitar la autenticación entre un cliente y un servidor. En este caso, el cliente es Redshift y el servidor es Amazon MSK. Para obtener más información acerca de cómo configurar la ingesta de streaming con mTLS, consulte Autenticación con mTLS para la ingesta de transmisiones de Redshift desde orígenes de Apache Kafka.

AUTHENTICATION_ARN

El ARN del certificado AWS Certificate Manager utilizado por Amazon Redshift para la autenticación mtls con Amazon MSK. El ARN está disponible en la consola de ACM al elegir el certificado emitido.

CLUSTER_ARN

Para la ingesta de streaming, CLUSTER_ARN es el identificador del clúster de Amazon Managed Streaming para Apache Kafka desde el que está transmitiendo. Al utilizar CLUSTER_ARN, se requiere una política de rol de IAM que incluya el permiso `kafka:GetBootstrapBrokers`. Esta opción se proporciona por compatibilidad con versiones anteriores. Actualmente, recomendamos utilizar la opción de URI de bootstrap-broker para conectarse a clústeres de Amazon Managed Streaming para Apache Kafka. Para obtener más información, consulte Ingesta de streaming.

## Notas de uso

Para conocer los límites de uso del catálogo de datos de Athena, consulte Límites de Athena en la Referencia general de AWS.

Para conocer los límites cuando se utiliza el AWS Glue Data Catalog, consulte Límites de AWS Glue en la Referencia general de AWS.

Estos límites no se aplican a un metaalmacén Hive.

Hay un máximo de 9 900 esquemas por base de datos. Para obtener más información, consulte Cuotas y límites en la Guía de administración de Amazon Redshift.

Para anular el registro del esquema, utilice el comando DROP SCHEMA.

Para ver detalles de los esquemas externos, consulte las siguientes vistas del sistema:

- SVV_EXTERNAL_SCHEMAS
- SVV_EXTERNAL_TABLES
- SVV_EXTERNAL_COLUMNS

## Ejemplos

El siguiente ejemplo crea un esquema externo utilizando una base de datos en un catálogo de datos denominado `sampledb` en la región Oeste de EE. UU. (Oregón). Utilice este ejemplo con un catálogo de datos de Athena o AWS Glue.

```
create external schema spectrum_schema
from data catalog
database 'sampledb'
region 'us-west-2'
iam_role 'arn:aws:iam::123456789012:role/MySpectrumRole';
```

En el siguiente ejemplo, se crean un esquema externo y una nueva base de datos externa denominada `spectrum_db`.

```
create external schema spectrum_schema
from data catalog
database 'spectrum_db'
iam_role 'arn:aws:iam::123456789012:role/MySpectrumRole'
create external database if not exists;
```

En el siguiente ejemplo, se crea un esquema externo a través de una base de datos de metaalmacén Hive denominada `hive_db`.

```
create external schema hive_schema
from hive metastore
database 'hive_db'
uri '172.10.10.10' port 99
iam_role 'arn:aws:iam::123456789012:role/MySpectrumRole';
```

En el siguiente ejemplo, se encadenan roles con el fin de utilizar el rol `myS3Role` para acceder a Amazon S3 y se utiliza `myAthenaRole` para acceder al catálogo de datos. Para obtener más información, consulte Encadenamiento de roles de IAM en Amazon Redshift Spectrum.

```
create external schema spectrum_schema
from data catalog
database 'spectrum_db'
iam_role 'arn:aws:iam::123456789012:role/myRedshiftRole,arn:aws:iam::123456789012:role/
myS3Role'
catalog_role 'arn:aws:iam::123456789012:role/myAthenaRole'
```

```
create external database if not exists;
```

En el siguiente ejemplo, se crea un esquema externo que referencia una base de datos de Aurora PostgreSQL.

```
CREATE EXTERNAL SCHEMA [IF NOT EXISTS] myRedshiftSchema
FROM POSTGRES
DATABASE 'my_aurora_db' SCHEMA 'my_aurora_schema'
URI 'endpoint to aurora hostname' PORT 5432
IAM_ROLE 'arn:aws:iam::123456789012:role/MyAuroraRole'
SECRET_ARN 'arn:aws:secretsmanager:us-east-2:123456789012:secret:development/
MyTestDatabase-AbCdEf'
```

En el siguiente ejemplo, se crea un esquema externo para referenciar la base de datos sales_db importada en el clúster consumidor.

```
CREATE EXTERNAL SCHEMA sales_schema FROM REDSHIFT DATABASE 'sales_db' SCHEMA 'public';
```

En el siguiente ejemplo, se crea un esquema externo que referencia una base de datos de Aurora MySQL.

```
CREATE EXTERNAL SCHEMA [IF NOT EXISTS] myRedshiftSchema
FROM MYSQL
DATABASE 'my_aurora_db'
URI 'endpoint to aurora hostname'
IAM_ROLE 'arn:aws:iam::123456789012:role/MyAuroraRole'
SECRET_ARN 'arn:aws:secretsmanager:us-east-2:123456789012:secret:development/
MyTestDatabase-AbCdEf'
```

## CREATE EXTERNAL TABLE

Crea una nueva tabla externa en el esquema especificado. Todas las tablas externas deben crearse en un esquema externo. La ruta de búsqueda no es compatible con esquemas externos y tablas externas. Para obtener más información, consulte CREATE EXTERNAL SCHEMA.

Además de las tablas externas creadas con el comando CREATE EXTERNAL TABLE, Amazon Redshift puede referencia tablas externas definidas en un catálogo de AWS Glue o AWS Lake Formation, o bien, en un metastore de Apache Hive. Utilice el comando CREATE EXTERNAL SCHEMA para registrar una base de datos externa definida en el catálogo externo y para hacer que

las tablas externas estén disponibles para usarse en Amazon Redshift. Si la tabla externa ya existe en un catálogo de AWS Glue o AWS Lake Formation o en un metastore de Hive, no necesita crear la tabla con CREATE EXTERNAL TABLE. Para ver las tablas externas, consulte la vista del sistema SVV_EXTERNAL_TABLES.

Si ejecuta el comando CREATE EXTERNAL TABLE AS, puede crear una tabla externa basada en la definición de la columna de una consulta y escribir los resultados de esa consulta en Amazon S3. Los resultados están en Apache Parquet o formato de texto delimitado. Si la tabla externa tiene una clave o claves de partición, Amazon Redshift particiona los archivos nuevos según esas claves de partición y registra las particiones nuevas en el catálogo externo de forma automática. Para obtener más información acerca de CREATE EXTERNAL TABLE AS, consulte Notas de uso.

Puede consultar una tabla externa mediante la misma sintaxis SELECT que utiliza con otras tablas de Amazon Redshift. También puede utilizar la sintaxis INSERT para escribir archivos nuevos en la ubicación de la tabla externa en Amazon S3. Para obtener más información, consulte INSERT (tabla externa).

Para crear una vista con una tabla externa, incluya la cláusula WITH NO SCHEMA BINDING en la instrucción CREATE VIEW.

No puede ejecutar CREATE EXTERNAL TABLE en una transacción (BEGIN … END). Para obtener más información acerca de las transacciones, consulte Aislamiento serializable.

## Privilegios necesarios

Para crear tablas externas, debe ser el propietario del esquema externo o un superusuario. Para transferir la propiedad de un esquema externo, use ALTER SCHEMA para cambiar el propietario. El acceso a tablas externas está controlado por el acceso al esquema externo. No puede usar permisos GRANT o REVOKE en una tabla externa. En su lugar, conceda o revoque USAGE en el esquema externo.

Las Notas de uso tienen información adicional sobre permisos específicos para tablas externas.

## Sintaxis

```
CREATE EXTERNAL TABLE
external_schema.table_name
(column_name data_type [, …] )
[ PARTITIONED BY (col_name data_type [, … ] )]
```

```
[ { ROW FORMAT DELIMITED row_format |
  ROW FORMAT SERDE 'serde_name'
  [ WITH SERDEPROPERTIES ( 'property_name' = 'property_value' [, ...] ) ] } ]
STORED AS file_format
LOCATION { 's3://bucket/folder/' | 's3://bucket/manifest_file' }
[ TABLE PROPERTIES ( 'property_name'='property_value' [, ...] ) ]
```

A continuación, se muestra la sintaxis de CREATE EXTERNAL TABLE AS.

```
CREATE EXTERNAL TABLE
external_schema.table_name
[ PARTITIONED BY (col_name [, … ] ) ]
[ ROW FORMAT DELIMITED row_format ]
STORED AS file_format
LOCATION { 's3://bucket/folder/' }
[ TABLE PROPERTIES ( 'property_name'='property_value' [, ...] ) ]
 AS
 { select_statement }
```

## Parámetros

external_schema.table_name

El nombre de la tabla que se creará, clasificada por un nombre de esquema externo. Las tablas externas deben crearse en un esquema externo. Para obtener más información, consulte CREATE EXTERNAL SCHEMA.

La longitud máxima del nombre de la tabla es de 127 bytes; los nombres más largos se truncan en 127 bytes. Puede usar caracteres multibyte UTF-8 de hasta un máximo de cuatro bytes. Amazon Redshift establece un límite de 9900 tablas por clúster, incluidas las tablas temporales definidas por el usuario y las tablas temporales creadas por Amazon Redshift durante el procesamiento de consultas o el mantenimiento del sistema. De manera opcional, puede completar el nombre de la tabla con el nombre de base de datos. En el siguiente ejemplo, el nombre de base de datos es `spectrum_db`, el nombre de esquema externo es `spectrum_schema` y el nombre de tabla es `test`.

```
create external table spectrum_db.spectrum_schema.test (c1 int)
stored as parquet
location 's3://amzn-s3-demo-bucket/myfolder/';
```

Si no existe la base de datos o el esquema especificados, no se crea la tabla y la instrucción devuelve un error. No puede crear tablas o vistas en las bases de datos del sistema `template0`, `template1`, `padb_harvest` o `sys:internal`.

El nombre de la tabla debe ser un nombre único para el esquema especificado.

Para obtener más información acerca de los nombres válidos, consulte Nombres e identificadores.

( column_name data_type )

El nombre y el tipo de datos de cada columna que se crea.

La longitud máxima del nombre de la columna es de 127 bytes; los nombres más largos se truncan en 127 bytes. Puede usar caracteres multibyte UTF-8 de hasta un máximo de cuatro bytes. No puede especificar nombres de columna `"$path"` o `"$size"`. Para obtener más información acerca de los nombres válidos, consulte Nombres e identificadores.

De manera predeterminada, Amazon Redshift crea tablas externas con las pseudocolumnas `$path` y `$size`. Puede deshabilitar la creación de pseudocolumnas para una sesión estableciendo el parámetro de configuración `spectrum_enable_pseudo_columns` en `false`. Para obtener más información, consulte Pseudocolumnas .

Si las pseudocolumnas están habilitadas, el número máximo de columnas que se pueden definir en una única tabla es 1 598. Si las pseudocolumnas no están habilitadas, el número máximo de columnas que se pueden definir en una única tabla es de 1600.

Si va a crear una "tabla ancha", asegúrese de que la lista de columnas no supere los límites de ancho de las filas para los resultados intermedios durante la carga y el procesamiento de consultas. Para obtener más información, consulte Notas de uso.

Para un comando CREATE EXTERNAL TABLE AS, no es necesaria una lista de columnas, ya que las columnas se obtienen de la consulta.

data_type

Se admiten los siguientes Tipos de datos:

- SMALLINT (INT2)
- INTEGER (INT, INT4)
- BIGINT (INT8)
- DECIMAL (NUMERIC)

- REAL (FLOAT4)
- DOUBLE PRECISION (FLOAT8)
- BOOLEAN (BOOL)
- CHAR (CHARACTER)
- VARCHAR (CHARACTER VARYING)
- VARBYTE (CHARACTER VARYING): se puede usar con archivos de datos Parquet y ORC, y solo con tablas que no sean de partición.
- DATE: solo se puede utilizar con archivos de datos de texto, Parquet u ORC, o como una columna de partición.
- TIMESTAMP

Para DATE, puede utilizar los formatos tal y como se describe a continuación. Para los valores de mes representados mediante dígitos, se admiten los siguientes formatos:

- `mm-dd-yyyy`, por ejemplo, `05-01-2017`. Esta es la opción predeterminada.
- `yyyy-mm-dd`, donde el año está representado por más de 2 dígitos. Por ejemplo, `2017-05-01`.

Para los valores de mes representados mediante abreviaturas de tres letras, se admiten los siguientes formatos:

- `mmm-dd-yyyy`, por ejemplo, `may-01-2017`. Esta es la opción predeterminada.
- `dd-mmm-yyyy`, donde el año está representado por más de 2 dígitos. Por ejemplo, `01-may-2017`.
- `yyyy-mmm-dd`, donde el año está representado por más de 2 dígitos. Por ejemplo, `2017-may-01`.

En el caso de los valores de año sistemáticamente inferiores a 100, el año se calcula de la siguiente manera:

- Si el año es inferior a 70, el año se calcula como el año más 2000. Por ejemplo, la fecha 05-01-17 en el formato `mm-dd-yyyy` se convierte en `05-01-2017`.
- Si el año es inferior a 100 o mayor que 69, el año se calcula como el año más 1900. Por ejemplo, la fecha 05-01-89 en el formato `mm-dd-yyyy` se convierte en `05-01-1989`.
- Para los valores de año representados por dos dígitos, agregue ceros al principio para representar el año en 4 dígitos.

Los valores de marca temporal de los archivos de texto deben tener el formato `yyyy-mm-dd HH:mm:ss.SSSSSS`, tal y como se muestra en el siguiente valor de marca temporal: `2017-05-01 11:30:59.000000`.

La longitud de una columna VARCHAR se expresa en bytes no en caracteres. Por ejemplo, una columna VARCHAR(12) puede contener 12 caracteres de un byte o 6 caracteres de dos bytes. Cuando se realiza una consulta a una tabla externa, los resultados se truncan para ajustar el tamaño de columna sin devolver un error. Para obtener más información, consulte Almacenamiento y rangos.

Para obtener un rendimiento óptimo, le recomendamos que especifique el menor tamaño de columna que se adapte a sus datos. Para encontrar el tamaño máximo en bytes para los valores de una columna, use la función OCTET_LENGTH. El siguiente ejemplo devuelve el tamaño máximo de valores en la columna de correo electrónico.

```
select max(octet_length(email)) from users;

max
---
 62
```

PARTITIONED BY (col_name data_type [, … ] )

Una cláusula que define una tabla particionada con una o más columnas de partición. Se usa un directorio de datos separado para cada combinación especificada, lo que puede mejorar el rendimiento de la consulta en algunas circunstancias. Las columnas particionadas no existen dentro de los propios datos de la tabla. Recibirá un error si utiliza un valor para col_name (nombre_de_columna) igual al de una columna de la tabla.

Después de crear una tabla con particiones, modifique la tabla mediante una instrucción ALTER TABLE … ADD PARTITION para registrar nuevas particiones en el catálogo externo. Cuando agrega una partición, define la ubicación de la subcarpeta en Amazon S3 que contiene los datos de partición.

Por ejemplo, si la tabla `spectrum.lineitem_part` se define con `PARTITIONED BY (l_shipdate date)`, ejecute el siguiente comando ALTER TABLE para agregar una partición.

```
ALTER TABLE spectrum.lineitem_part ADD PARTITION (l_shipdate='1992-01-29')
LOCATION 's3://spectrum-public/lineitem_partition/l_shipdate=1992-01-29';
```

Si utiliza CREATE EXTERNAL TABLE AS, no tendrá que ejecutar ALTER TABLE...ADD PARTITION. Amazon Redshift registra de forma automática las particiones nuevas en el catálogo externo. Amazon Redshift también escribe de forma automática los datos correspondientes en las particiones de Amazon S3 en función de la clave o las claves de partición definidas en la tabla.

Para ver particiones, consulte la vista del sistema SVV_EXTERNAL_PARTITIONS.

> **Note**
>
> Para un comando CREATE EXTERNAL TABLE AS, no es necesario especificar el tipo de datos de la columna de partición porque esta columna se obtiene de la consulta.

ROW FORMAT DELIMITED rowformat (formato_de_fila)

Una cláusula que especifica el formato de los datos subyacentes. Los valores posibles para rowformat (formato_de_fila) son los siguientes:

- LINES TERMINATED BY 'delimiter'
- FIELDS TERMINATED BY 'delimiter'

Especifique un solo carácter ASCII para 'delimiter (delimitador)'. Puede especificar caracteres ASCII no imprimibles mediante octales, con el formato '*ddd*', donde *d* es un dígito octal (0-7) hasta “\177”. El siguiente ejemplo especifica el carácter BEL (campana) usando la representación octal.

```
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\007'
```

Si se omite ROW FORMAT DELIMITED, el formato predeterminado es DELIMITED FIELDS TERMINATED BY '\A' (inicio de encabezado) y LINES TERMINATED BY '\n' (nueva línea).

ROW FORMAT SERDE 'serde_name (nombre_de_serde)' , [WITH SERDEPROPERTIES ( 'property_name' = 'property_value' [, ...] ) ]

Una cláusula que especifica el formato SERDE para los datos subyacentes.

'serde_name (nombre_de_serde)'

El nombre del SerDe. Puede especificar los siguientes formatos:

- org.apache.hadoop.hive.serde2.RegexSerDe
- com.amazonaws.glue.serde.GrokSerDe

- org.apache.hadoop.hive.serde2.OpenCSVSerde

  Este parámetro admite la siguiente propiedad SerDE para OpenCSvSerde:

  ```
  'wholeFile' = 'true'
  ```

  Configure la propiedad `wholeFile` como `true` para analizar correctamente los nuevos caracteres de línea (\n) dentro de las cadenas citadas para las solicitudes OpenCSV.
- org.openx.data.jsonserde.JsonSerDe
  - El SERDE JSON también admite archivos Ion.
  - El JSON debe tener un formato correcto.
  - Las marcas temporales en Ion y JSON deben usar el formato ISO8601.
  - Este parámetro admite la siguiente propiedad SerDE para JsonSerDe:

    ```
    'strip.outer.array'='true'
    ```

    Procesa archivos Ion/JSON que contienen una matriz muy grande encerrada en corchetes exteriores ([ … ]) como si contiene múltiples registros JSON dentro de la matriz.
- com.amazon.ionhiveserde.IonHiveSerDe

  El formato Amazon ION proporciona formatos de texto y binarios, además de tipos de datos. Para una tabla externa que hace referencia a datos en formato ION, debe asignar cada columna en la tabla externa al elemento correspondiente de los datos de formato ION. Para obtener más información, consulte Amazon Ion. También debe especificar los formatos de entrada y salida.

WITH SERDEPROPERTIES ( 'property_name' = 'property_value' [, ...] ) ]

De manera opcional, especifique los nombres y valores de las propiedades, separados por comas.

Si se omite ROW FORMAT DELIMITED, el formato predeterminado es DELIMITED FIELDS TERMINATED BY '\A' (inicio de encabezado) y LINES TERMINATED BY '\n' (nueva línea).

STORED AS file_format

El formato de archivo para los archivos de datos.

Los formatos válidos son los siguientes:

- PARQUET
- RCFILE (para datos que usan ColumnarSerDe únicamente, no LazyBinaryColumnarSerDe)
- SEQUENCEFILE
- TEXTFILE (para archivos de texto, incluidos los archivos JSON).
- ORC
- AVRO
- INPUTFORMAT 'input_format_classname' OUTPUTFORMAT 'output_format_classname'

El comando CREATE EXTERNAL TABLE AS solo admite dos formatos de archivo, TEXTFILE y PARQUET.

Para INPUTFORMAT y OUTPUTFORMAT, especifique un nombre de la clase, como se muestra en el siguiente ejemplo:

```
'org.apache.hadoop.mapred.TextInputFormat'
```

LOCATION { 's3://bucket/folder/' | 's3://bucket/manifest_file'}

Se trata de la ruta a la carpeta o al bucket de Amazon S3 que contiene los archivos de datos o un archivo de manifiesto donde se incluye una lista de rutas de objetos de Amazon S3. Los buckets deben estar en la misma región de AWS que el clúster de Amazon Redshift. Para obtener una lista de las regiones de AWS admitidas, consulte Limitaciones de Amazon Redshift Spectrum.

Si la ruta especifica un bucket o una carpeta, por ejemplo, `'s3://amzn-s3-demo-bucket/custdata/'`, Redshift Spectrum analiza los archivos en el bucket o la carpeta especificados, además de todas las subcarpetas. Redshift Spectrum omite los archivos ocultos y los archivos que comienzan con un carácter de subrayado.

Si la ruta especifica un archivo de manifiesto, el argumento `'s3://bucket/manifest_file'` debe referencia de forma explícita un solo archivo; por ejemplo, `'s3://amzn-s3-demo-bucket/manifest.txt'`. No puede hacer referencia a un prefijo de clave.

El manifiesto es un archivo de texto en formato JSON que muestra la dirección URL de cada archivo que se va a cargar desde Amazon S3 y el tamaño del archivo, en bytes. El URL incluye el nombre del bucket y la ruta de objeto completa para el archivo. Los archivos que se especifican en el manifiesto pueden estar en buckets diferentes, pero todos los buckets deben estar en la misma región de AWS que el clúster de Amazon Redshift. Si un archivo aparece dos veces, este

se carga dos veces. En el siguiente ejemplo, se muestra el JSON para un manifiesto que carga tres archivos.

```
{
  "entries": [
    {"url":"s3://amzn-s3-demo-bucket1/custdata.1", "meta": { "content_length":
 5956875 } },
    {"url":"s3://amzn-s3-demo-bucket1/custdata.2", "meta": { "content_length":
 5997091 } },
    {"url":"s3://amzn-s3-demo-bucket2/custdata.1", "meta": { "content_length":
 5978675 } }
  ]
}
```

Puede hacer que la inclusión de un archivo en particular sea obligatoria. Para ello, incluya una opción `mandatory` en el nivel de archivo en el manifiesto. Cuando consulta una tabla externa con un archivo obligatorio faltante, se produce un error en la instrucción SELECT. Asegúrese de que todos los archivos incluidos en la definición de la tabla externa están presentes. Si no están todos presentes, aparece un error que muestra el primer archivo obligatorio que no se encuentra. En el siguiente ejemplo, se muestra el JSON de un manifiesto con la opción `mandatory` establecida en `true`.

```
{
  "entries": [
    {"url":"s3://amzn-s3-demo-bucket1/custdata.1", "mandatory":true, "meta":
 { "content_length": 5956875 } },
    {"url":"s3://amzn-s3-demo-bucket1/custdata.2", "mandatory":false, "meta":
 { "content_length": 5997091 } },
    {"url":"s3://amzn-s3-demo-bucket2/custdata.1", "meta": { "content_length":
 5978675 } }
  ]
}
```

Para hacer referencia a los archivos creados mediante UNLOAD, puede usar el manifiesto creado mediante UNLOAD con el parámetro MANIFEST. El archivo de manifiesto es compatible con un archivo de manifiesto de COPY de Amazon S3, pero usa claves diferentes. Las claves que no se usan se omiten.

TABLE PROPERTIES ( 'property_name'='property_value' [, ...] )

Una cláusula que establece la definición de tabla de las propiedades de la tabla.

> **Note**
>
> Las propiedades de tabla distinguen entre mayúsculas y minúsculas.

'compression_type'='value (valor)'

Propiedad que establece el tipo de compresión que se va a utilizar si el nombre de archivo no contiene una extensión. Si establece esta propiedad y hay una extensión de archivo, la extensión se omite y se usa el valor establecido por la propiedad. Los valores válidos para el tipo de compresión son los siguientes:

- bzip2
- gzip
- none
- snappy

“data_cleansing_enabled”=“true/false”

Esta propiedad establece si el control de datos está activado para la tabla. Cuando “data_cleansing_enabled” se establece en true, el control de datos está activado para la tabla. Cuando “data_cleansing_enabled” se establece en false, el control de datos está desactivado para la tabla. A continuación, se muestra una lista de las propiedades de control de datos en el nivel de la tabla que controla esta propiedad:

- column_count_mismatch_handling
- invalid_char_handling
- numeric_overflow_handling
- replacement_char
- surplus_char_handling

Para ver ejemplos, consulta Ejemplos de control de datos.

“invalid_char_handling”=“valor”

Especifica la acción que se debe realizar cuando los resultados de la consulta contienen valores de caracteres UTF-8 no válidos. Puede especificar las siguientes acciones:

DISABLED

No lleva a cabo un control de caracteres no válido.

FAIL

Cancela las consultas que devuelven datos que contienen valores UTF-8 no válidos.

SET_TO_NULL

Reemplaza los valores UTF-8 no válidos por valores nulos.

DROP_ROW

Reemplaza cada valor de la fila por un valor nulo.

REPLACE

Reemplaza el carácter no válido por el carácter de reemplazo que se especifique mediante `replacement_char`.

“replacement_char”=“carácter”

Especifica el carácter de reemplazo que se va a utilizar al establecer `invalid_char_handling` en `REPLACE`.

“numeric_overflow_handling”=“valor”

Especifica la acción que se debe realizar cuando los datos de ORC contienen un entero (por ejemplo, BIGINT o int64) que es mayor que la definición de columna (por ejemplo, SMALLINT o int16). Puede especificar las siguientes acciones:

DISABLED

El control de caracteres no válidos está desactivado.

FAIL

Cancela la consulta cuando los datos incluyen caracteres no válidos.

SET_TO_NULL

Establece caracteres no válidos en valores nulos.

DROP_ROW

Establece cada valor de la fila como un valor nulo.

'surplus_bytes_handling'='valor'

Especifica cómo controlar los datos que se cargan que superan la longitud del tipo de datos definido para las columnas que contienen datos VARBYTE. De manera predeterminada, Redshift Spectrum establece el valor como nulo para los datos que superan el ancho de la columna.

Puede especificar las siguientes acciones que se realizarán cuando la consulta devuelva datos que superen la longitud del tipo de datos:

SET_TO_NULL

Reemplaza los datos que superan el ancho de la columna por valores nulos.

DISABLED

No lleva a cabo un control de bytes sobrantes.

FAIL

Cancela las consultas que devuelven datos que superan el ancho de la columna.

DROP_ROW

Elimina todas las filas que contienen datos que superan la anchura de columna.

TRUNCATE

Elimina los caracteres que superan el número máximo de caracteres definidos para la columna.

“surplus_char_handling”=“valor”

Especifica cómo controlar los datos que se cargan que superan la longitud del tipo de datos definido para las columnas que contienen datos VARCHAR, CHAR o cadenas. De manera predeterminada, Redshift Spectrum establece el valor como nulo para los datos que superan el ancho de la columna.

Puede especificar las siguientes acciones que se deben realizar cuando la consulta devuelva datos que superan el ancho de la columna:

SET_TO_NULL

Reemplaza los datos que superan el ancho de la columna por valores nulos.

DISABLED

No lleva a cabo un control de caracteres sobrantes.

FAIL

Cancela las consultas que devuelven datos que superan el ancho de la columna.

DROP_ROW

Reemplaza cada valor de la fila por un valor nulo.

TRUNCATE

Elimina los caracteres que superan el número máximo de caracteres definidos para la columna.

'column_count_mismatch_handling'='value’

Identifica si el archivo contiene menos o más valores para una fila que el número de columnas especificado en la definición de tabla externa. Esta propiedad solo está disponible para un formato de archivo de texto sin comprimir. Puede especificar las siguientes acciones:

DISABLED

El control de discrepancias en el recuento de columnas está desactivado.

FAIL

No se puede realizar la consulta si se detecta una discrepancia en el recuento de columnas.

SET_TO_NULL

Rellene los valores que faltan con NULL e ignore los valores adicionales de cada fila.

DROP_ROW

Elimine del análisis todas las filas que contengan un error de discrepancia en el recuento de columnas.

'numRows'='row_count'

Una propiedad que establece el valor de numRows para la definición de la tabla. Para actualizar de forma explícita las estadísticas de una tabla externa, establezca la propiedad numRows de manera que indique el tamaño de la tabla. Amazon Redshift no analiza las tablas externas para generar las estadísticas de las tablas que el optimizador de consultas emplea a la hora de crear un plan de consulta. Si no se configuran las estadísticas de tabla para una tabla externa, Amazon Redshift genera un plan de ejecución de consulta basado en un supuesto de que las tablas externas son las más grandes y las tablas locales son las más pequeñas.

'skip.header.line.count'='line_count'

Una propiedad que establece el número de filas que se omiten al principio de cada archivo de código fuente.

'serialization.null.format'=' '

Una propiedad que especifica que Spectrum debe devolver un valor NULL cuando hay una coincidencia exacta con el texto introducido en un campo.

'orc.schema.resolution'='mapping_type' (tipo_de_asignación)

Propiedad que establece el tipo de asignación de columnas en tablas que usan el formato de datos de ORC. Esta propiedad se omite con otros formatos de datos.

Los valores válidos para el tipo de asignación de columnas son los siguientes:

- nombre
- position

Si la propiedad orc.schema.resolution se omite, las columnas se asignan por nombre de manera predeterminada. Si orc.schema.resolution se establece en un valor que no es 'name' ni 'position', las columnas se asignan por posición. Para obtener más información sobre la asignación de columnas, consulte Asignación de columnas de tablas externas a columnas ORC.

> **Note**
>
> El comando COPY realiza asignaciones en archivos de datos de ORC únicamente por posición. La propiedad de tabla orc.schema.resolution no tiene ningún efecto en el comportamiento del comando COPY.

'write.parallel'='on / off'

Una propiedad que establece si CREATE EXTERNAL TABLE AS debe escribir datos en paralelo. De manera predeterminada, CREATE EXTERNAL TABLE AS escribe datos en paralelo en varios archivos, según el número de sectores en el clúster. La opción predeterminada está activada. Cuando 'write.parallel' está desactivado, CREATE EXTERNAL TABLE AS escribe en uno o más archivos de datos en serie en Amazon S3. Esta propiedad de tabla también se aplica a cualquier instrucción INSERT posterior en la misma tabla externa.

'write.maxfilesize.mb'='size'

Se trata de una propiedad que establece el tamaño máximo (en MB) de cada archivo escrito en Amazon S3 con CREATE EXTERNAL TABLE AS. El tamaño debe ser un entero válido

entre 5 y 6200. El tamaño máximo de archivo predeterminado es 6200 MB. Esta propiedad de tabla también se aplica a cualquier instrucción INSERT posterior en la misma tabla externa.

'write.kms.key.id'='value'

Puede especificar una clave AWS Key Management Service para habilitar el cifrado del lado del servidor (SSE) para objetos de Amazon S3, donde value es una de las opciones siguientes:

- `auto` para utilizar la clave de AWS KMS predeterminada almacenada en el bucket de Amazon S3
- clave de kms especficada para cifrar los datos

select_statement

Una instrucción que inserta una o más filas en la tabla externa definiendo cualquier consulta. Todas las filas que genera la consulta se escriben en Amazon S3 en formato de texto o Parquet según la definición de tabla.

## Ejemplos

Encontrará una selección de ejemplos en Ejemplos.

## Notas de uso

Este tema contiene notas de uso para CREATE EXTERNAL TABLE. No puede ver los detalles de las tablas de Amazon Redshift Spectrum con los mismos recursos que utiliza para las tablas de Amazon Redshift estándar, como PG_TABLE_DEF, STV_TBL_PERM, PG_CLASS o information_schema. Si su herramienta de análisis o inteligencia empresarial no reconoce las tablas externas de Redshift Spectrum, configure la aplicación para consultar SVV_EXTERNAL_TABLES y SVV_EXTERNAL_COLUMNS.

CREATE EXTERNAL TABLE AS

En algunos casos, puede ejecutar el comando CREATE EXTERNAL TABLE AS en un catálogo de datos de AWS Glue, un catálogo externo de AWS Lake Formation o un metastore de Apache Hive. En tales casos, se utiliza un rol de AWS Identity and Access Management (IAM) para crear el esquema externo. Este rol de IAM debe tener permisos de lectura y escritura en Amazon S3.

Si utiliza un catálogo de Lake Formation, el rol de IAM debe tener permiso para crear tablas en el catálogo. En este caso, también debe tener el permiso de ubicación del lago de datos en la ruta de

destino de Amazon S3. Este rol de IAM se convierte en el propietario de la nueva tabla AWS Lake Formation.

Para asegurarse de que los nombres de los archivos son únicos, Amazon Redshift utiliza el siguiente formato para el nombre de cada archivo cargado en Amazon S3 de manera predeterminada.

*<date>*_*<time>*_*<microseconds>*_*<query_id>*_*<slice-number>*_part_*<part-number>*.*<format>*.

Un ejemplo es `20200303_004509_810669_1007_0001_part_00.parquet`.

Tenga en cuenta lo siguiente al ejecutar el comando CREATE EXTERNAL TABLE AS:

- La ubicación de Amazon S3 debe estar vacía.
- Amazon Redshift solo admite los formatos PARQUET y TEXTFILE cuando se utiliza la cláusula STORED AS.
- No es necesario definir una lista de definición de columna. Los nombres de columna y los tipos de datos de columna de la nueva tabla externa se obtienen directamente de la consulta SELECT.
- No es necesario definir el tipo de datos de la columna de partición en la cláusula PARTITIONED BY. Si especifica una clave de partición, el nombre de esta columna debe existir en el resultado de la consulta SELECT. Cuando tiene varias columnas de partición, no importa el orden en la consulta SELECT. Amazon Redshift utiliza el orden definido en la cláusula PARTITIONED BY para crear la tabla externa.
- Amazon Redshift particiona de forma automática los archivos de salida en carpetas de partición en función de los valores de la clave de partición. De manera predeterminada, Amazon Redshift quita las columnas de partición de los archivos de salida.
- No se admite la cláusula LINES TERMINATED BY 'delimitador'.
- No se admite la cláusula ROW FORMAT SERDE 'serde_name'.
- No se admite el uso de archivos de manifiesto. Por lo tanto, no se puede definir la cláusula LOCATION en un archivo de manifiesto en Amazon S3.
- Amazon Redshift actualiza de forma automática la propiedad de la tabla 'numRows' al final del comando.
- La propiedad de tabla 'compression_type' sólo acepta 'none' o 'snappy' para el formato de archivo PARQUET.
- Amazon Redshift no admite la cláusula LIMIT en la consulta SELECT externa. En su lugar, puede utilizar una cláusula LIMIT anidada.

- Puede utilizar STL_UNLOAD_LOG para realizar un seguimiento de los archivos que cada operación CREATE EXTERNAL TABLE AS escribe en Amazon S3.

### Permisos necesarios para crear y consultar tablas externas

Para crear tablas externas, asegúrese de que es el propietario del esquema externo o un superusuario. Para transferir la propiedad de un esquema externo, utilice ALTER SCHEMA. En el siguiente ejemplo, se modifica el propietario del esquema `spectrum_schema` a `newowner`.

```
alter schema spectrum_schema owner to newowner;
```

Para ejecutar una consulta de Redshift Spectrum, necesita los siguientes permisos:

- Permiso de uso para el esquema
- Permiso para crear tablas temporales en la base de datos actual

En el siguiente ejemplo, se concede permiso de uso para el esquema `spectrum_schema` al grupo de usuarios `spectrumusers`.

```
grant usage on schema spectrum_schema to group spectrumusers;
```

En el siguiente ejemplo, se concede un permiso temporal para la base de datos `spectrumdb` al grupo de usuarios `spectrumusers`.

```
grant temp on database spectrumdb to group spectrumusers;
```

### Pseudocolumnas

De manera predeterminada, Amazon Redshift crea tablas externas con las pseudocolumnas $path y $size. Seleccione estas columnas para ver la ruta a los archivos de datos en Amazon S3 y el tamaño de los archivos de datos de cada fila devuelta por una consulta. Los nombres de las columnas $path y $size deben estar delimitados con comillas dobles. Las cláusulas SELECT * no devuelven las pseudocolumnas. Debe incluir explícitamente los nombres de columna $path y $size en la consulta, tal y como se muestra en el siguiente ejemplo.

```
select "$path", "$size"
from spectrum.sales_part
where saledate = '2008-12-01';
```

Puede deshabilitar la creación de pseudocolumnas en una sesión estableciendo el parámetro de configuración spectrum_enable_pseudo_columns en false.

> **Important**
>
> La selección de $size o $path genera cargos, ya que Redshift Spectrum analiza los archivos de datos en Amazon S3 para determinar el tamaño del conjunto de resultados. Para obtener más información, consulte Precios de Amazon Redshift.

Configuración de opciones de control de datos

Puede establecer parámetros de tabla para especificar el control de entrada de los datos que se consultan en tablas externas, incluidos los siguientes:

- Caracteres sobrantes en columnas que contengan datos de VARCHAR, CHAR y cadenas. Para obtener más información, consulte la propiedad de tabla externa `surplus_char_handling`.
- Caracteres no válidos en columnas que contengan datos de VARCHAR, CHAR y cadenas. Para obtener más información, consulte la propiedad de tabla externa `invalid_char_handling`.
- Carácter de reemplazo que se utilizará al especificar REPLACE para la propiedad de tabla externa `invalid_char_handling`.
- Control de desbordamiento de conversión en columnas que contengan datos enteros y decimales. Para obtener más información, consulte la propiedad de tabla externa `numeric_overflow_handling`.
- Surplus_bytes_handling para especificar la gestión de entrada de los bytes sobrantes en las columnas que contienen datos varbyte. Para obtener más información, consulte la propiedad de tabla externa `surplus_bytes_handling`.

## Ejemplos

En el siguiente ejemplo, se crea una tabla denominada SALES en el esquema externo `spectrum` de Amazon Redshift. Los datos están en archivos de texto delimitados por tabulaciones. La cláusula TABLE PROPERTIES configura la propiedad numRows con 170 000 filas.

Según la identidad que utilice para ejecutar CREATE EXTERNAL TABLE, es posible que deba configurar permisos de IAM. Como práctica recomendada, aconsejamos asociar las políticas de

permisos a un rol de IAM y luego asignarlo a los usuarios y grupos según sea necesario. Para obtener más información, consulte Administración de identidades y accesos en Amazon Redshift.

```
create external table spectrum.sales(
salesid integer,
listid integer,
sellerid integer,
buyerid integer,
eventid integer,
saledate date,
qtysold smallint,
pricepaid decimal(8,2),
commission decimal(8,2),
saletime timestamp)
row format delimited
fields terminated by '\t'
stored as textfile
location 's3://redshift-downloads/tickit/spectrum/sales/'
table properties ('numRows'='170000');
```

En el siguiente ejemplo, se crea una tabla que utiliza JsonSerDe para hacer referencia a datos con formato JSON.

```
create external table spectrum.cloudtrail_json (
event_version int,
event_id bigint,
event_time timestamp,
event_type varchar(10),
awsregion varchar(20),
event_name varchar(max),
event_source varchar(max),
requesttime timestamp,
useragent varchar(max),
recipientaccountid bigint)
row format serde 'org.openx.data.jsonserde.JsonSerDe'
with serdeproperties (
'dots.in.keys' = 'true',
'mapping.requesttime' = 'requesttimestamp'
) location 's3://amzn-s3-demo-bucket/json/cloudtrail';
```

En el siguiente ejemplo CREATE EXTERNAL TABLE AS se crea una tabla externa no particionada. Luego, escribe el resultado de la consulta SELECT con Apache Parquet en la ubicación de destino de Amazon S3.

```
CREATE EXTERNAL TABLE spectrum.lineitem
STORED AS parquet
LOCATION 'S3://amzn-s3-demo-bucket/cetas/lineitem/'
AS SELECT * FROM local_lineitem;
```

En el siguiente ejemplo, se crea una tabla externa particionada y se incluyen las columnas de partición en la consulta SELECT.

```
CREATE EXTERNAL TABLE spectrum.partitioned_lineitem
PARTITIONED BY (l_shipdate, l_shipmode)
STORED AS parquet
LOCATION 'S3://amzn-s3-demo-bucket/cetas/partitioned_lineitem/'
AS SELECT l_orderkey, l_shipmode, l_shipdate, l_partkey FROM local_table;
```

Para obtener una lista de bases de datos existentes en el catálogo de datos externos, consulte la vista del sistema SVV_EXTERNAL_DATABASES.

```
select eskind,databasename,esoptions from svv_external_databases order by databasename;
```

```
eskind | databasename | esoptions
-------+--------------
+-------------------------------------------------------------------------------
     1 | default      | {"REGION":"us-
west-2","IAM_ROLE":"arn:aws:iam::123456789012:role/mySpectrumRole"}
     1 | sampledb     | {"REGION":"us-
west-2","IAM_ROLE":"arn:aws:iam::123456789012:role/mySpectrumRole"}
     1 | spectrumdb   | {"REGION":"us-
west-2","IAM_ROLE":"arn:aws:iam::123456789012:role/mySpectrumRole"}
```

Para ver detalles de tablas externas, consulte las vistas del sistema SVV_EXTERNAL_TABLES y SVV_EXTERNAL_COLUMNS.

En el siguiente ejemplo, se consulta la vista SVV_EXTERNAL_TABLES.

```
select schemaname, tablename, location from svv_external_tables;
```

```
schemaname | tablename            | location
-----------+----------------------
+-----------------------------------------------------
spectrum   | sales                | s3://redshift-downloads/tickit/spectrum/sales
spectrum   | sales_part           | s3://redshift-downloads/tickit/spectrum/
sales_partition
```

En el siguiente ejemplo, se consulta la vista SVV_EXTERNAL_COLUMNS.

```
select * from svv_external_columns where schemaname like 'spectrum%' and tablename
 ='sales';
```

```
schemaname | tablename | columnname | external_type | columnnum | part_key
-----------+-----------+------------+---------------+-----------+---------
spectrum   | sales     | salesid    | int           |         1 |        0
spectrum   | sales     | listid     | int           |         2 |        0
spectrum   | sales     | sellerid   | int           |         3 |        0
spectrum   | sales     | buyerid    | int           |         4 |        0
spectrum   | sales     | eventid    | int           |         5 |        0
spectrum   | sales     | saledate   | date          |         6 |        0
spectrum   | sales     | qtysold    | smallint      |         7 |        0
spectrum   | sales     | pricepaid  | decimal(8,2)  |         8 |        0
spectrum   | sales     | commission | decimal(8,2)  |         9 |        0
spectrum   | sales     | saletime   | timestamp     |        10 |        0
```

Para ver las particiones de la tabla, use la siguiente consulta.

```
select schemaname, tablename, values, location
from svv_external_partitions
where tablename = 'sales_part';
```

```
schemaname | tablename  | values         | location
-----------+------------+----------------
+-------------------------------------------------------------------------
spectrum   | sales_part | ["2008-01-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-01
spectrum   | sales_part | ["2008-02-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-02
spectrum   | sales_part | ["2008-03-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-03
```

```
spectrum   | sales_part | ["2008-04-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-04
spectrum   | sales_part | ["2008-05-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-05
spectrum   | sales_part | ["2008-06-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-06
spectrum   | sales_part | ["2008-07-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-07
spectrum   | sales_part | ["2008-08-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-08
spectrum   | sales_part | ["2008-09-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-09
spectrum   | sales_part | ["2008-10-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-10
spectrum   | sales_part | ["2008-11-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-11
spectrum   | sales_part | ["2008-12-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-12
```

El siguiente ejemplo devuelve el tamaño total de los archivos de datos relacionados de una tabla externa.

```
select distinct "$path", "$size"
   from spectrum.sales_part;

 $path                                                                      | $size
----------------------------------------------------------------------------+-------
s3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-01/ |  1616
s3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-02/ |  1444
s3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-02/ |  1444
```

### Ejemplos de partición

Para crear una tabla externa particionada por fecha, ejecute el siguiente comando.

```
create external table spectrum.sales_part(
salesid integer,
listid integer,
sellerid integer,
buyerid integer,
eventid integer,
dateid smallint,
qtysold smallint,
```

```
pricepaid decimal(8,2),
commission decimal(8,2),
saletime timestamp)
partitioned by (saledate date)
row format delimited
fields terminated by '|'
stored as textfile
location 's3://redshift-downloads/tickit/spectrum/sales_partition/'
table properties ('numRows'='170000');
```

Para agregar las particiones, ejecute los siguientes comandos ALTER TABLE.

```
alter table spectrum.sales_part
add if not exists partition (saledate='2008-01-01')
location 's3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-01/';
alter table spectrum.sales_part
add if not exists partition (saledate='2008-02-01')
location 's3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-02/';
alter table spectrum.sales_part
add if not exists partition (saledate='2008-03-01')
location 's3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-03/';
alter table spectrum.sales_part
add if not exists partition (saledate='2008-04-01')
location 's3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-04/';
alter table spectrum.sales_part
add if not exists partition (saledate='2008-05-01')
location 's3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-05/';
alter table spectrum.sales_part
add if not exists partition (saledate='2008-06-01')
location 's3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-06/';
alter table spectrum.sales_part
add if not exists partition (saledate='2008-07-01')
location 's3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-07/';
alter table spectrum.sales_part
add if not exists partition (saledate='2008-08-01')
location 's3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-08/';
alter table spectrum.sales_part
add if not exists partition (saledate='2008-09-01')
location 's3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-09/';
alter table spectrum.sales_part
add if not exists partition (saledate='2008-10-01')
location 's3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-10/';
alter table spectrum.sales_part
```

```
add if not exists partition (saledate='2008-11-01')
location 's3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-11/';
alter table spectrum.sales_part
add if not exists partition (saledate='2008-12-01')
location 's3://redshift-downloads/tickit/spectrum/sales_partition/saledate=2008-12/';
```

Para seleccionar los datos de la tabla particionada, ejecute la siguiente consulta.

```
select top 10 spectrum.sales_part.eventid, sum(spectrum.sales_part.pricepaid)
from spectrum.sales_part, event
where spectrum.sales_part.eventid = event.eventid
  and spectrum.sales_part.pricepaid > 30
  and saledate = '2008-12-01'
group by spectrum.sales_part.eventid
order by 2 desc;
```

```
eventid | sum
--------+---------
    914 | 36173.00
   5478 | 27303.00
   5061 | 26383.00
   4406 | 26252.00
   5324 | 24015.00
   1829 | 23911.00
   3601 | 23616.00
   3665 | 23214.00
   6069 | 22869.00
   5638 | 22551.00
```

Para ver particiones de la tabla externa, consulte la vista del sistema SVV_EXTERNAL_PARTITIONS.

```
select schemaname, tablename, values, location from svv_external_partitions
where tablename = 'sales_part';
```

```
schemaname | tablename  | values         | location
-----------+------------+----------------
+-------------------------------------------------
spectrum   | sales_part | ["2008-01-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-01
```

```
spectrum   | sales_part | ["2008-02-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-02
spectrum   | sales_part | ["2008-03-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-03
spectrum   | sales_part | ["2008-04-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-04
spectrum   | sales_part | ["2008-05-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-05
spectrum   | sales_part | ["2008-06-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-06
spectrum   | sales_part | ["2008-07-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-07
spectrum   | sales_part | ["2008-08-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-08
spectrum   | sales_part | ["2008-09-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-09
spectrum   | sales_part | ["2008-10-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-10
spectrum   | sales_part | ["2008-11-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-11
spectrum   | sales_part | ["2008-12-01"] | s3://redshift-downloads/tickit/spectrum/
sales_partition/saledate=2008-12
```

Ejemplos de formato de fila

A continuación, se muestra un ejemplo de cómo especificar los parámetros ROW FORMAT SERDE para archivos de datos almacenados con formato AVRO.

```
create external table spectrum.sales(salesid int, listid int, sellerid int,
 buyerid int, eventid int, dateid int, qtysold int, pricepaid decimal(8,2), comment
 VARCHAR(255))
ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.avro.AvroSerDe'
WITH SERDEPROPERTIES ('avro.schema.literal'='{\"namespace\": \"dory.sample\",\"name\":
 \"dory_avro\",\"type\": \"record\", \"fields\": [{\"name\":\"salesid\", \"type\":\"int
\"},
{\"name\":\"listid\", \"type\":\"int\"},
{\"name\":\"sellerid\", \"type\":\"int\"},
{\"name\":\"buyerid\", \"type\":\"int\"},
{\"name\":\"eventid\",\"type\":\"int\"},
{\"name\":\"dateid\",\"type\":\"int\"},
{\"name\":\"qtysold\",\"type\":\"int\"},
{\"name\":\"pricepaid\", \"type\": {\"type\": \"bytes\", \"logicalType\": \"decimal\",
 \"precision\": 8, \"scale\": 2}}, {\"name\":\"comment\",\"type\":\"string\"}]}')
```

```
STORED AS AVRO
location 's3://amzn-s3-demo-bucket/avro/sales' ;
```

A continuación se muestra un ejemplo de especificación de los parámetros ROW FORMAT SERDE mediante RegEx.

```
create external table spectrum.types(
cbigint bigint,
cbigint_null bigint,
cint int,
cint_null int)
row format serde 'org.apache.hadoop.hive.serde2.RegexSerDe'
with serdeproperties ('input.regex'='([^\\x01]+)\\x01([^\\x01]+)\\x01([^\\x01]+)\
\x01([^\\x01]+)')
stored as textfile
location 's3://amzn-s3-demo-bucket/regex/types';
```

A continuación se muestra un ejemplo de especificación de los parámetros ROW FORMAT SERDE mediante Grok.

```
create external table spectrum.grok_log(
timestamp varchar(255),
pid varchar(255),
loglevel varchar(255),
progname varchar(255),
message varchar(255))
row format serde 'com.amazonaws.glue.serde.GrokSerDe'
with serdeproperties ('input.format'='[DFEWI], \\[%{TIMESTAMP_ISO8601:timestamp} #
%{POSINT:pid:int}\\] *(?<loglevel>:DEBUG|FATAL|ERROR|WARN|INFO) -- +%{DATA:progname}:
 %{GREEDYDATA:message}')
stored as textfile
location 's3://DOC-EXAMPLE-BUCKET/grok/logs';
```

A continuación, se muestra un ejemplo en el que se define un registro de acceso al servidor de Amazon S3 en un bucket de S3. Puede utilizar Redshift Spectrum para consultar los registros de acceso de Amazon S3.

```
CREATE EXTERNAL TABLE spectrum.mybucket_s3_logs(
bucketowner varchar(255),
bucket varchar(255),
requestdatetime varchar(2000),
```

```
remoteip varchar(255),
requester varchar(255),
requested varchar(255),
operation varchar(255),
key varchar(255),
requesturi_operation varchar(255),
requesturi_key varchar(255),
requesturi_httpprotoversion varchar(255),
httpstatus varchar(255),
errorcode varchar(255),
bytessent bigint,
objectsize bigint,
totaltime varchar(255),
turnaroundtime varchar(255),
referrer varchar(255),
useragent varchar(255),
versionid varchar(255)
)
ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.RegexSerDe'
WITH SERDEPROPERTIES (
'input.regex' = '([^ ]*) ([^ ]*) \\[(.*?)\\] ([^ ]*) ([^ ]*) ([^ ]*) ([^ ]*) ([^ ]*)
 \"([^ ]*)\\s*([^ ]*)\\s*([^ ]*)\" (- |[^ ]*) ([^ ]*) ([^ ]*) ([^ ]*) ([^ ]*) ([^ ]*)
 ([^ ]*) (\"[^\"]*\") ([^ ]*).*$')
LOCATION 's3://amzn-s3-demo-bucket/s3logs’;
```

A continuación, se muestra un ejemplo de cómo especificar los parámetros ROW FORMAT SERDE para datos con formato ION.

```
CREATE EXTERNAL TABLE tbl_name (columns)
ROW FORMAT SERDE 'com.amazon.ionhiveserde.IonHiveSerDe'
STORED AS
INPUTFORMAT 'com.amazon.ionhiveserde.formats.IonInputFormat'
OUTPUTFORMAT 'com.amazon.ionhiveserde.formats.IonOutputFormat'
LOCATION 's3://amzn-s3-demo-bucket/prefix'
```

Ejemplos de control de datos

Los ejemplos siguientes acceden al archivo spi_global_rankings.csv. Puede cargar el archivo `spi_global_rankings.csv` en un bucket de Amazon S3 para probar estos ejemplos.

En el siguiente ejemplo, se crea el esquema externo `schema_spectrum_uddh` y la base de datos `spectrum_db_uddh`. En el caso de `aws-account-id`, ingrese su ID de cuenta de AWS y, para `role-name`, ingrese el nombre de rol de Redshift Spectrum.

```
create external schema schema_spectrum_uddh
from data catalog
database 'spectrum_db_uddh'
iam_role 'arn:aws:iam::aws-account-id:role/role-name'
create external database if not exists;
```

En el siguiente ejemplo, se crea una tabla externa `soccer_league` en el esquema externo `schema_spectrum_uddh`.

```
CREATE EXTERNAL TABLE schema_spectrum_uddh.soccer_league
(
  league_rank smallint,
  prev_rank   smallint,
  club_name   varchar(15),
  league_name varchar(20),
  league_off  decimal(6,2),
  league_def  decimal(6,2),
  league_spi  decimal(6,2),
  league_nspi integer
)
ROW FORMAT DELIMITED
    FIELDS TERMINATED BY ','
    LINES TERMINATED BY '\n\l'
stored as textfile
LOCATION 's3://spectrum-uddh/league/'
table properties ('skip.header.line.count'='1');
```

Verifique el número de filas en la tabla `soccer_league`.

```
select count(*) from schema_spectrum_uddh.soccer_league;
```

Se muestra el número de filas.

```
count
645
```

En la siguiente consulta, se muestran los 10 clubes principales. Dado que el club `Barcelona` tiene un carácter no válido en la cadena, se muestra un NULL en el nombre.

```
select league_rank,club_name,league_name,league_nspi
from schema_spectrum_uddh.soccer_league
where league_rank between 1 and 10;
```

```
league_rank club_name league_name   league_nspi
1  Manchester City Barclays Premier Lea  34595
2  Bayern Munich German Bundesliga  34151
3  Liverpool Barclays Premier Lea  33223
4  Chelsea  Barclays Premier Lea  32808
5  Ajax  Dutch Eredivisie  32790
6  Atletico  Madrid Spanish Primera Divi 31517
7  Real Madrid Spanish Primera Divi  31469
8  NULL         Spanish Primera Divi            31321
9  RB Leipzig German Bundesliga  31014
10  Paris Saint-Ger French Ligue 1   30929
```

En el siguiente ejemplo, se modifica la tabla `soccer_league` para especificar las propiedades de tabla externa `invalid_char_handling`, `replacement_char` y `data_cleansing_enabled` con objeto de insertar un signo de interrogación (?) como reemplazo de caracteres no previstos.

```
alter  table schema_spectrum_uddh.soccer_league
set table properties
 ('invalid_char_handling'='REPLACE','replacement_char'='?','data_cleansing_enabled'='true');
```

En el siguiente ejemplo, se consulta la tabla `soccer_league` para equipos con una clasificación del 1 al 10.

```
select league_rank,club_name,league_name,league_nspi
from schema_spectrum_uddh.soccer_league
where league_rank between 1 and 10;
```

Dado que las propiedades de la tabla se han modificado, los resultados muestran los 10 principales clubes, con el carácter de sustitución del signo de interrogación (?) en la octava fila para el club `Barcelona`.

```
league_rank club_name league_name  league_nspi
```

```
1  Manchester City Barclays Premier Lea 34595
2  Bayern Munich German Bundesliga 34151
3  Liverpool Barclays Premier Lea 33223
4  Chelsea  Barclays Premier Lea 32808
5  Ajax  Dutch Eredivisie 32790
6  Atletico Madrid Spanish Primera Divi 31517
7  Real Madrid Spanish Primera Divi 31469
8  Barcel?na Spanish Primera Divi 31321
9  RB Leipzig German Bundesliga 31014
10  Paris Saint-Ger French Ligue 1  30929
```

En el siguiente ejemplo, se modifica la tabla `soccer_league` para especificar las propiedades de tabla externa `invalid_char_handling` con objeto de eliminar filas con caracteres no previstos.

```
alter table schema_spectrum_uddh.soccer_league
set table properties
 ('invalid_char_handling'='DROP_ROW','data_cleansing_enabled'='true');
```

En el siguiente ejemplo, se consulta la tabla `soccer_league` para equipos con una clasificación del 1 al 10.

```
select league_rank,club_name,league_name,league_nspi
from schema_spectrum_uddh.soccer_league
where league_rank between 1 and 10;
```

En los resultados se muestran los clubes principales, pero no se incluye la octava fila, que correspondería al club `Barcelona`.

```
league_rank   club_name         league_name            league_nspi
1             Manchester City   Barclays Premier Lea   34595
2             Bayern Munich     German Bundesliga      34151
3             Liverpool         Barclays Premier Lea   33223
4             Chelsea           Barclays Premier Lea   32808
5             Ajax              Dutch Eredivisie       32790
6             Atletico Madrid   Spanish Primera Divi   31517
7             Real Madrid       Spanish Primera Divi   31469
9             RB Leipzig        German Bundesliga      31014
10            Paris Saint-Ger   French Ligue 1         30929
```

# CREATE EXTERNAL VIEW

La característica de versión preliminar de las vistas del catálogo de datos solo está disponible en las siguientes regiones.

- Este de EE. UU. (Ohio) (us-east-2)
- Este de EE. UU. (Norte de Virginia) (us-east-1)
- EE. UU. Oeste (Norte de California) (us-west-1)
- Asia-Pacífico (Tokio) (ap-northeast-1)
- Europa (Irlanda) (eu-west-1)
- Europa (Estocolmo) (eu-north-1)

Crea una vista de catálogo de datos. Las vistas de catálogo de datos son un esquema de vista única que funciona con otros motores de SQL, como Amazon Athena y Amazon EMR. Puede consultar la vista desde el motor que elija. Para obtener más información acerca de las vistas del catálogo de datos, consulte Creación de vistas del catálogo de datos.

## Sintaxis

```
CREATE EXTERNAL VIEW schema_name.view_name [ IF NOT EXISTS ]
{catalog_name.schema_name.view_name | awsdatacatalog.dbname.view_name |
 external_schema_name.view_name}
AS query_definition;
```

## Parámetros

schema_name.view_name

Es el esquema asociado a la base de datos de AWS Glue, seguido del nombre de la vista.

PROTECTED

Especifica que el comando CREATE EXTERNAL VIEW solo debería completarse si la consulta incluida en la query_definition se puede completar correctamente.

IF NOT EXISTS

Crea la vista si aún no existe.

catalog_name.schema_name.view_name | awsdatacatalog.dbname.view_name | external_schema_name.view_name

Es la notación del esquema que se utilizará al crear la vista. Puede especificar el uso de AWS Glue Data Catalog, una base de datos de Glue que haya creado o un esquema externo que haya creado. Consulte CREATE DATABASE y CREATE EXTERNAL SCHEMA para obtener más información.

query_definition

Es la definición de la consulta SQL que Amazon Redshift ejecuta para modificar la vista.

### Ejemplos

El siguiente ejemplo crea una vista del catálogo de datos denominada sample_schema.glue_data_catalog_view.

```
CREATE EXTERNAL PROTECTED VIEW sample_schema.glue_data_catalog_view IF NOT EXISTS
AS SELECT * FROM sample_database.remote_table "remote-table-name";
```

## CREATE FUNCTION

Crea un nueva función escalar definida por el usuario (UDF) mediante una cláusula SQL SELECT o un programa Python.

Para obtener más información y ejemplos, consulta Funciones definidas por el usuario en Amazon Redshift.

### Privilegios necesarios

Debe haber obtenido permiso de una de las siguientes maneras para ejecutar CREATE OR REPLACE FUNCTION:

- Para CREATE FUNCTION:
  - Un superusuario puede utilizar lenguajes de confianza y que no sean de confianza para crear funciones.
  - Los usuarios con el privilegio CREATE [ OR REPLACE ] FUNCTION pueden crear funciones con lenguajes de confianza.

- Para REPLACE FUNCTION:
  - Superusuario
  - Usuarios con el privilegio CREATE [ OR REPLACE ] FUNCTION
  - Propietario de la función

## Sintaxis

```
CREATE [ OR REPLACE ] FUNCTION f_function_name
( { [py_arg_name  py_arg_data_type |
sql_arg_data_type } [ , ... ] ] )
RETURNS data_type
{ VOLATILE | STABLE | IMMUTABLE }
AS $$
  { python_program | SELECT_clause }
$$ LANGUAGE { plpythonu | sql }
```

## Parámetros

OR REPLACE

Especifica si ya existe una función con el mismo nombre y los mismos tipos de datos de argumento de entrada o firma que este, debe reemplazarse la función existente. Solo puede reemplazar una función con una nueva función que defina un conjunto idéntico de tipos de datos. Debe ser un superusuario para reemplazar una función.

Si define una función con el mismo nombre que una función existente, pero con una firma diferente, crea una nueva función. En otras palabras, se sobrecarga el nombre de la función. Para obtener más información, consulte Sobrecarga de los nombres de función.

f_function_name

El nombre de la función. Si especifica un nombre de esquema (como `myschema.myfunction`), la función se crea con el esquema especificado. De lo contrario, la función se crea en el esquema actual. Para obtener más información acerca de los nombres válidos, consulte Nombres e identificadores.

Recomendamos que utilice el prefijo `f_` en los nombres de todas las UDF. Amazon Redshift reserva el prefijo `f_` para los nombres de las UDF; por lo tanto, cuando utiliza el prefijo `f_`, se asegura de que el nombre de la UDF no entre en conflicto con el nombre de una función

SQL integrada en Amazon Redshift existente o futura. Para obtener más información, consulte Prevención de conflictos al dar nombre a las UDF.

Puede definir más de una función con el mismo nombre de función si los tipos de datos de los argumentos de entrada son diferentes. En otras palabras, se sobrecarga el nombre de la función. Para obtener más información, consulte Sobrecarga de los nombres de función.

py_arg_name py_arg_data_type | sql_arg_data_type

Para una UDF de Python, una lista de nombres de argumento de entrada y tipos de datos. Para una UDF de SQL, una lista tipos de datos, sin nombres de argumento. En una UDF de Python, haga referencia a los argumentos usando los nombres de argumento. En una UDF de SQL, haga referencia a los argumentos usando $1, $2, etc., en función del orden de los argumentos en la lista de argumentos.

Para una UDF de SQL, los tipos de datos de entrada y salida pueden ser cualquier tipo de datos estándar de Amazon Redshift. Para una UDF Python, los tipos de datos de entrada y salida pueden ser SMALLINT, INTEGER, BIGINT, DECIMAL, REAL, DOUBLE PRECISION, BOOLEAN, CHAR, VARCHAR, DATE o TIMESTAMP. Además, las funciones definidas por el usuario (UDF) de Python admiten un tipo de datos de ANYELEMENT. Este tipo de datos se convierte automáticamente en un tipo de datos estándar basado en el tipo de datos correspondiente al argumento proporcionado en tiempo de ejecución. Si varios argumentos utilizan ANYELEMENT, todos se resuelven con el mismo tipo de datos durante el tiempo de ejecución, según el primer argumento ANYELEMENT de la lista. Para obtener más información, consulte Tipos de datos de UDF de Python y Tipos de datos.

Puede especificar un máximo de 32 argumentos.

RETURNS data_type

El tipo de datos del valor que la función devuelve. El tipo de datos RETURNS puede ser cualquier tipo de datos estándar de Amazon Redshift. Además, las UDF de Python pueden usar un tipo de datos de ANYELEMENT, que se convierte automáticamente en un tipo de datos estándar basado en el argumento proporcionado en tiempo de ejecución. Si especifica ANYELEMENT para el tipo de datos que se devuelve, al menos un argumento debe usar ANYELEMENT. El tipo de datos que se devuelve coincide con el tipo de datos proporcionado por el argumento ANYELEMENT cuando se convoca la función. Para obtener más información, consulte Tipos de datos de UDF de Python.

VOLATILE | STABLE | IMMUTABLE

Informa al optimizador de consultas acerca de la volatilidad de la función.

Obtendrá la mejor optimización si etiqueta la función con la categoría de volatilidad válida más estricta. No obstante, si la categoría es demasiado estricta, hay un riesgo de que el optimizador omita por error algunas ejecuciones, lo que resultaría en un conjunto de resultados incorrectos. Las categorías de volatilidad, de la menos estricta a la más estricta, son las siguientes:

- VOLATILE
- STABLE
- IMMUTABLE

VOLATILE

Dados los mismos argumentos, la función puede devolver resultados diferentes en ejecuciones consecutivas, incluso para las filas de una única instrucción. El optimizador de consultas no puede realizar ninguna suposición acerca del comportamiento de una función volátil, por lo que una consulta que utiliza una función volátil debe reevaluar la función para cada fila de entrada.

STABLE

Dados los mismos argumentos, se garantiza que la función devuelve los mismos resultados para todas las filas procesadas dentro de una única instrucción. La función puede devolver resultados diferentes cuando se evoca en diferentes instrucciones. Esta categoría permite que el optimizador mejore diversas ejecuciones de la función dentro de una única instrucción a una única ejecución para la instrucción.

IMMUTABLE

Dados los mismos argumentos, la función siempre devuelve el mismo resultado. Cuando una consulta evoca una función `IMMUTABLE` con argumentos constantes, el optimizador evalúa previamente la función.

AS $$ statement $$

Una construcción que contiene la instrucción que se ejecutará. Las palabras claves literales `AS $$` y $$ son obligatorias.

En Amazon Redshift, la instrucción debe incluirse en la función con un formato llamado entrecomillado con símbolo de dólar. Cualquier elemento que se encuentre dentro de los signos se transmite exactamente como es. No necesita incluir en una secuencia de escape los caracteres especiales porque el contenido de la cadena se escribe literalmente.

Con el entrecomillado de dólar, se utilizan un par de signos de dólar ($$) para representar el inicio y el final de la instrucción que se va a ejecutar, tal y como se muestra en el siguiente ejemplo.

```
$$ my statement $$
```

De manera opcional, entre los signos de dólar de cada par, puede especificar una cadena para ayudar a identificar la instrucción. La cadena que usa debe ser la misma en el inicio y en el final de los pares de signos. La cadena distingue entre mayúsculas y minúsculas, y sigue las mismas restricciones que un identificador sin comillas, excepto que no puede contener signos de dólar. En el siguiente ejemplo, se usa la cadena `test`.

```
$test$ my statement $test$
```

Para obtener más información acerca de los signos de dólar, consulte Dollar-quoted String Constants bajo Lexical Structure en la documentación de PostgreSQL.

python_program

Un programa Python ejecutable válido que devuelve un valor. La instrucción que pase con la función debe ajustarse a los requisitos de sangría, tal y como se especifica en la guía de estilo para el código de Python del sitio web de Python. Para obtener más información, consulte Compatibilidad del lenguaje Python con las UDF.

SQL_clause

Una cláusula SQL SELECT.

La cláusula SELECT no puede incluir ninguno de los tipos de cláusulas siguientes:

- FROM
- INTO
- WHERE
- GROUP BY
- ORDER BY
- LIMIT

LANGUAGE { plpythonu | sql }

Para Python, especifique `plpythonu`. Para SQL, especifique `sql`. Debe tener permiso para usar el lenguaje de SQL o plpythonu. Para obtener más información, consulte Seguridad y permisos de UDF.

## Notas de uso

### Funciones anidadas

Puede llamar a otra función definida por el usuario (UDF) de SQL desde una UDF de SQL. La función anidada debe existir cuando se ejecuta el comando CREATE FUNCTION. Amazon Redshift no realiza un seguimiento de las dependencias de las UDF; por lo tanto, si se elimina la función anidada, Amazon Redshift no devuelve ningún error. No obstante, la UDF producirá un error si la función anidada no existe. Por ejemplo, la siguiente función llama a la función `f_sql_greater` en la cláusula SELECT.

```
create function f_sql_commission (float, float )
  returns float
stable
as $$
  select f_sql_greater ($1, $2)
$$ language sql;
```

### Privilegios y seguridad de las UDF

Para crear una UDF, debe tener permiso para usar el lenguaje de SQL o plpythonu (Python). De manera predeterminada, se concede el permiso USAGE ON LANGUAGE SQL a PUBLIC. No obstante, debe conceder explícitamente el permiso de USAGE ON LANGUAGE PLPYTHONU a usuarios o grupos específicos.

Para revocar los permisos de uso de SQL, primero revoque el uso de PUBLIC. A continuación, conceda los permisos de uso de SQL a solo los usuarios o grupos que tienen permiso para crear UDF de SQL. En el siguiente ejemplo, se revocan los permisos de uso de SQL de PUBLIC y después se conceden permisos de uso al grupo de usuarios `udf_devs`.

```
revoke usage on language sql from PUBLIC;
grant usage on language sql to group udf_devs;
```

Para ejecutar una UDF, debe tener permisos de ejecución para cada función. De manera predeterminada, el permiso de ejecución de nuevas UDF está establecido en PUBLIC. Para restringir el uso, revoque el permiso de ejecución PUBLIC para la función. A continuación, conceda el privilegio a usuarios o grupos específicos.

En el siguiente ejemplo, se revocan los permisos de ejecución en la función `f_py_greater` de PUBLIC y después se conceden permisos de uso al grupo de usuarios `udf_devs`.

```
revoke execute on function f_py_greater(a float, b float) from PUBLIC;
grant execute on function f_py_greater(a float, b float) to group udf_devs;
```

De manera predeterminada, los superusuarios tienen todos los privilegios.

Para obtener más información, consulte GRANT y REVOKE.

## Ejemplos

### Ejemplo de una UDF escalar de Python

En el siguiente ejemplo, se crea una UDF de Python que compara dos enteros y devuelve el valor mayor.

```
create function f_py_greater (a float, b float)
  returns float
stable
as $$
  if a > b:
    return a
  return b
$$ language plpythonu;
```

En el siguiente ejemplo, se consulta la tabla SALES y se evoca la nueva función `f_py_greater` para devolver COMMISSION o 20 por ciento de PRICEPAID, el que represente el mayor valor.

```
select f_py_greater (commission, pricepaid*0.20) from sales;
```

### Ejemplo de una UDF escalar de SQL

En el siguiente ejemplo, se crea una función que compara dos números y devuelve el valor más grande.

```
create function f_sql_greater (float, float)
  returns float
stable
as $$
  select case when $1 > $2 then $1
    else $2
  end
$$ language sql;
```

La siguiente consulta llama a la nueva función `f_sql_greater` para consultar la tabla SALES y devuelve COMMISSION o el 20 por ciento de PRICEPAID, lo que sea mayor.

```
select f_sql_greater (commission, pricepaid*0.20) from sales;
```

# CREATE GROUP

Define un nuevo grupo de usuarios. Solo un superusuario puede crear un grupo.

## Sintaxis

```
CREATE GROUP group_name
[ [ WITH ] [ USER username ] [, ...] ]
```

## Parámetros

group_name

Nombre del nuevo grupo de usuarios. Los nombres de grupos que comienzan con dos guiones bajos se reservan para el uso interno de Amazon Redshift. Para obtener más información acerca de los nombres válidos, consulte Nombres e identificadores.

WITH

Sintaxis opcional para indicar parámetros adicionales para CREATE GROUP.

USER

Agregue uno o más usuarios al grupo.

username

Nombre del usuario que desea agregar al grupo.

## Ejemplos

En el siguiente ejemplo, se crea un grupo de usuarios llamado ADMIN_GROUP con dos usuarios: ADMIN1 y ADMIN2.

```
create group admin_group with user admin1, admin2;
```

# CREATE IDENTITY PROVIDER

Define un nuevo proveedor de identidades. Solo un superusuario puede crear un proveedor de identidades.

## Sintaxis

```
CREATE IDENTITY PROVIDER identity_provider_name TYPE type_name
NAMESPACE namespace_name
[PARAMETERS parameter_string]
[APPLICATION_ARN arn]
[IAM_ROLE iam_role]
[AUTO_CREATE_ROLES
    [ TRUE [ { INCLUDE | EXCLUDE } GROUPS LIKE filter_pattern] |
      FALSE
    ]
  ];
```

## Parámetros

identity_provider_name

Nombre del nuevo proveedor de identidades. Para obtener más información acerca de los nombres válidos, consulte Nombres e identificadores.

type_name

El proveedor de identidades con el que interactuar. En la actualidad, Azure y AWSIDC son los únicos proveedores de identidad admitidos.

namespace_name

El espacio de nombres. Se trata de un identificador abreviado único para el directorio del proveedor de identidades.

parameter_string

Una cadena que contiene un objeto JSON con el formato correcto que contiene los parámetros y valores necesarios para el proveedor de identidades específico.

arn

El nombre de recurso de Amazon (ARN) de una aplicación administrada por el IAM Identity Center. Este parámetro solo se aplica cuando el tipo de proveedor de identidades es AWSIDC.

iam_role

El rol de IAM que proporciona permisos para realizar la conexión al IAM Identity Center. Este parámetro solo se aplica cuando el tipo de proveedor de identidades es AWSIDC.

auto_create_roles

Habilita o deshabilita la característica de creación automática de roles. Si el valor es TRUE, Amazon Redshift habilita la característica de creación automática de roles. Si el valor es FALSE, Amazon Redshift desactiva la característica de creación automática de roles. Si no se especifica el valor de este parámetro, Amazon Redshift determina el valor mediante la siguiente lógica:

- Si se proporciona `AUTO_CREATE_ROLES` pero no se especifica el valor, el valor se establece en TRUE.
- Si no se proporciona `AUTO_CREATE_ROLES` y el proveedor de identidad es AWSIDC, el valor se establece en FALSE.
- Si no se proporciona `AUTO_CREATE_ROLES` y el proveedor de identidad es Azure, el valor se establece en TRUE.

Para incluir grupos, especifique `INCLUDE`. El valor predeterminado es vacío, lo que significa incluir todos los grupos si `AUTO_CREATE_ROLES` está activado.

Para excluir grupos, especifique `EXCLUDE`. El valor predeterminado es vacío, lo que significa que no se excluirá ningún grupo si `AUTO_CREATE_ROLES` está activado.

filter_pattern

Una expresión de caracteres UTF-8 válida con un patrón para hacer coincidir los nombres de grupo. La opción LIKE realiza una coincidencia que distingue entre mayúsculas y minúsculas y admite los siguientes metacaracteres de coincidencia de patrones:

| Metacarácter | Descripción |
|---|---|
| `%` | Coincide con cualquier secuencia de cero o más caracteres. |
| `_` | Coincide con cualquier carácter. |

Si filter_pattern no contiene metacaracteres, solo representa la propia cadena; en ese caso, LIKE actúa igual que el operador de igualdad.

filter_pattern admite los siguientes caracteres:

- Caracteres alfabéticos en mayúsculas y minúsculas (A-Z y a-z)
- Números (0-9)
- Los siguientes caracteres especiales:

```
_ % ^ * + ? { } , $
```

## Ejemplos

En el ejemplo siguiente se crea un proveedor de identidades denominado oauth_standard, con un TIPO azure, para establecer la comunicación con Microsoft Azure Active Directory (AD).

```
CREATE IDENTITY PROVIDER oauth_standard TYPE azure
NAMESPACE 'aad'
PARAMETERS '{"issuer":"https://sts.windows.net/2sdfdsf-d475-420d-b5ac-667adad7c702/",
"client_id":"87f4aa26-78b7-410e-bf29-57b39929ef9a",
"client_secret":"BUAH~ewrqewrqwerUUY^%tHe1oNZShoiU7",
"audience":["https://analysis.windows.net/powerbi/connector/AmazonRedshift"]
}'
```

Puede conectar una aplicación administrada de IAM Identity Center con un clúster aprovisionado existente o un grupo de trabajo de Amazon Redshift sin servidor. Esto le permite administrar el acceso a una base de datos de Redshift a través del IAM Identity Center. Para ello, ejecute un comando de SQL como el siguiente ejemplo. Debe ser administrador de bases de datos.

```
CREATE IDENTITY PROVIDER "redshift-idc-app" TYPE AWSIDC
NAMESPACE 'awsidc'
APPLICATION_ARN 'arn:aws:sso::123456789012:application/ssoins-12345f67fe123d4/apl-
a0b0a12dc123b1a4'
IAM_ROLE 'arn:aws:iam::123456789012:role/MyRedshiftRole';
```

En este caso, el ARN de la aplicación identifica la aplicación administrada a la que conectarse. Puede encontrarla ejecutando `SELECT * FROM SVV_IDENTITY_PROVIDERS;`.

Para obtener más información sobre el uso de CREATE IDENTITY PROVIDER, incluidos ejemplos adicionales, consulte Federación de proveedores de identidades (IdP) nativos para Amazon Redshift. Para obtener más información sobre la configuración de una conexión de IAM Identity Center desde Redshift, consulte Conectar Redshift con IAM Identity Center para ofrecer a los usuarios una experiencia de inicio de sesión único.

# CREATE LIBRARY

Instala una biblioteca de Python, que esta disponible para que los usuarios la incorporen cuando crean una función definida por el usuario (UDF) con el comando CREATE FUNCTION. El tamaño total de las bibliotecas instaladas por el usuario no puede superar los 100 MB.

CREATE LIBRARY no puede ejecutarse dentro de un bloque de transacción (BEGIN … END). Para obtener más información acerca de las transacciones, consulte Aislamiento serializable.

Amazon Redshift es compatible con la versión 2.7 de Python. Para obtener más información, consulte www.python.org.

Para obtener más información, consulte Ejemplo: Importación de módulos personalizados desde la biblioteca de Python.

## Privilegios necesarios

Los siguientes privilegios son necesarios para CREATE LIBRARY:

- Superusuario
- Usuarios con el privilegio CREATE LIBRARY o con el privilegio del lenguaje especificado

## Sintaxis

```
CREATE [ OR REPLACE ] LIBRARY library_name LANGUAGE plpythonu
FROM
{ 'https://file_url'
| 's3://bucketname/file_name'
authorization
  [ REGION [AS] 'aws_region']
  IAM_ROLE { default | ‘arn:aws:iam::<Cuenta de AWS-id>:role/<role-name>’ }
}
```

## Parámetros

### OR REPLACE

Especifica que se reemplaza la biblioteca existente si ya existe una biblioteca con el mismo nombre. REPLACE se confirma inmediatamente. Si una UDF que depende de la biblioteca se ejecuta simultáneamente, es posible que la UDF falle o devuelva resultados inesperados, incluso

si la UDF se ejecuta dentro de una transacción. Debe ser el propietario o un superusuario para reemplazar una biblioteca.

library_name

El nombre de la biblioteca que se instalará. No se puede crear una biblioteca que contenga un módulo con el mismo nombre que un módulo de la Biblioteca estándar de Python o un módulo de Python preinstalado en Amazon Redshift. Si una biblioteca existente instalada por el usuario utiliza el mismo paquete de Python que la biblioteca que se instalará, debe eliminar la biblioteca existente antes de instalar la biblioteca nueva. Para obtener más información, consulte Compatibilidad del lenguaje Python con las UDF.

LANGUAGE plpythonu

El lenguaje que se utilizará. Python (plpythonu) es el único lenguaje compatible. Amazon Redshift es compatible con la versión 2.7 de Python. Para obtener más información, consulte www.python.org.

FROM

La ubicación del archivo de biblioteca. Puede especificar un nombre de objeto y bucket de Amazon S3, o bien, puede especificar una dirección URL para descargar el archivo desde un sitio web público. La biblioteca debe comprimirse en la forma de un archivo `.zip`. Para obtener más información, consulte Building and Installing Python Modules en la documentación de Python.

https://file_url

El URL para descargar el archivo desde un sitio web público. El URL puede contener hasta tres redireccionamientos. A continuación se muestra un ejemplo del URL de un archivo.

```
'https://www.example.com/pylib.zip'
```

s3://bucket_name/file_name

Se trata de la ruta a un objeto único de Amazon S3 que contiene el archivo de la biblioteca. En el siguiente ejemplo, se muestra una ruta de objeto de Amazon S3.

```
's3://amzn-s3-demo-bucket/my-pylib.zip'
```

Si especifica un bucket de Amazon S3, también debe proporcionar credenciales para un usuario de AWS que tenga permiso para descargar el archivo.

> **Important**
>
> Si el bucket de Amazon S3 no se encuentra en la misma región de AWS que el clúster de Amazon Redshift, debe utilizar la opción REGION para especificar la región de AWS en la que se encuentran los datos. El valor de aws_region debe coincidir con una región de AWS enumerada en la tabla en la descripción del parámetro REGION para el comando COPY.

authorization

Se trata de una cláusula que inicia el método que el clúster utiliza para la autenticación y la autorización con objeto de acceder al bucket de Amazon S3 que contiene el archivo de biblioteca. El clúster debe tener permisos para obtener acceso a Amazon S3 con las acciones LIST y GET.

La sintaxis de la autorización es la misma que para la autorización del comando COPY. Para obtener más información, consulte Parámetros de autorización.

```
IAM_ROLE { default | 'arn:aws:iam::<Cuenta de AWS-id>:role/<role-name>'
```

Utilice la palabra clave predeterminada para que Amazon Redshift utilice el rol de IAM configurado como predeterminado y asociado al clúster cuando se ejecuta el comando CREATE LIBRARY.

Utilice el nombre de recurso de Amazon (ARN), de un rol de IAM que el clúster utiliza para la autenticación y la autorización. Si especifica IAM_ROLE, no puede utilizar ACCESS_KEY_ID, SECRET_ACCESS_KEY, SESSION_TOKEN ni CREDENTIALS.

De manera opcional, si el bucket de Amazon S3 utiliza cifrado de servidor, proporcione la clave de cifrado en la cadena credentials-args. Si utiliza credenciales de seguridad temporales, proporcione el token temporal en la cadena credentials-args.

Para obtener más información, consulte Credenciales de seguridad temporales.

REGION [AS] aws_region

Se trata de la región de AWS donde está ubicado el bucket de Amazon S3. REGION es necesario cuando el bucket de Amazon S3 no está en la misma región de AWS que el clúster de Amazon Redshift. El valor de aws_region debe coincidir con una región de AWS enumerada en la tabla en la descripción del parámetro REGION para el comando COPY.

De manera predeterminada, CREATE LIBRARY asume que el bucket de Amazon S3 está ubicado en la misma región de AWS que el clúster de Amazon Redshift.

### Ejemplos

Los siguientes dos ejemplos instalan el módulo de Python urlparse, que está comprimido en un archivo llamado `urlparse3-1.0.3.zip`.

El siguiente comando instala una biblioteca de UDF denominada `f_urlparse` desde un paquete que se ha cargado en un bucket de Amazon S3 ubicado en la región EE. UU. Este.

```
create library f_urlparse
language plpythonu
from 's3://amzn-s3-demo-bucket/urlparse3-1.0.3.zip'
credentials 'aws_iam_role=arn:aws:iam::<aws-account-id>:role/<role-name>'
region as 'us-east-1';
```

En el siguiente ejemplo, se instala una biblioteca denominada `f_urlparse` desde un archivo de biblioteca en un sitio web.

```
create library f_urlparse
language plpythonu
from 'https://example.com/packages/urlparse3-1.0.3.zip';
```

## CREATE MASKING POLICY

Crea una nueva política de enmascaramiento dinámico de datos para ofuscar los datos de un determinado formato. Para obtener más información sobre el enmascaramiento dinámico de datos, consulte Enmascaramiento de datos dinámico.

Los superusuarios y los usuarios o roles que tienen el rol sys:secadmin pueden crear una política de enmascaramiento.

### Sintaxis

```
CREATE MASKING POLICY
   policy_name [IF NOT EXISTS]
   WITH (input_columns)
```

```
    USING (masking_expression);
```

### Parámetros

policy_name

Nombre de la política de enmascaramiento. La política de enmascaramiento no puede tener el mismo nombre que otra política de enmascaramiento que ya existe en la base de datos.

input_columns

Una tupla de nombres de columnas con el formato (tipo de col1, tipo de col2, ...).

Los nombres de columna se utilizan como entrada de la expresión de enmascaramiento. Los nombres de columna no tienen que coincidir con los nombres de columna que se enmascaran, pero los tipos de datos de entrada y de salida sí deben coincidir.

masking_expression

Expresión SQL que se utiliza para transformar las columnas de destino. Se puede escribir mediante funciones de manipulación de datos, como las funciones de manipulación de cadenas, o junto con funciones definidas por el usuario escritas en SQL, Python o con AWS Lambda. Puede incluir una tupla de expresiones de columna para las políticas de enmascaramiento que tienen múltiples salidas. Si usa una constante como expresión de enmascaramiento, debe convertirla explícitamente a un tipo que coincida con el tipo de entrada.

Debe tener el permiso USAGE en todas las funciones definidas por el usuario que utilice en la expresión de enmascaramiento.

## CREATE MATERIALIZED VIEW

Crea una vista materializada basada en una o más tablas de Amazon Redshift. También puede basar las vistas materializadas en tablas externas creadas mediante Spectrum o una consulta federada. Para obtener información acerca de Spectrum, consulte Amazon Redshift Spectrum. Para obtener información acerca de las consultas federadas, consulte Consulta de datos con consultas federadas en Amazon Redshift.

### Sintaxis

```
CREATE MATERIALIZED VIEW mv_name
```

```
[ BACKUP { YES | NO } ]
[ table_attributes ]
[ AUTO REFRESH { YES | NO } ]
AS query
```

## Parámetros

BACKUP

Una cláusula que especifica si la vista materializada debe incluirse en instantáneas de clústeres manuales y automáticos.

Para las vistas materializadas que no contienen datos críticos, especifique BACKUP NO para ahorrar tiempo de procesamiento al momento de crear instantáneas y restaurar contenido a partir de ellas, y para reducir espacio de almacenamiento en Amazon Simple Storage Service. El ajuste BACKUP NO no afecta la replicación automática de datos a otros nodos dentro del clúster, por lo que las vistas materializadas con BACKUP NO especificado se restauran en el caso de un error del nodo. El valor predeterminado es BACKUP YES.

table_attributes

Cláusula que especifica cómo se distribuyen los datos de la vista materializada, incluida la siguiente:

- El estilo de distribución para la vista materializada en formato `DISTSTYLE { EVEN | ALL | KEY }`. Si omite esta cláusula, el estilo de distribución es `EVEN`. Para obtener más información, consulte Estilos de distribución.
- La clave de distribución de la vista materializada en formato `DISTKEY ( distkey_identifier )`. Para obtener más información, consulte Designación de estilos de distribución.
- Clave de ordenación de la vista materializada en formato `SORTKEY ( column_name [, ...] )`. Para obtener más información, consulte Claves de clasificación.

AS query

Se trata de una instrucción `SELECT` válida que define la vista materializada y su contenido. El conjunto de resultados de la consulta define las columnas y filas de la vista materializada. Para obtener información sobre las limitaciones al crear vistas materializadas, consulte Limitaciones.

Además, las construcciones específicas del lenguaje SQL utilizadas en la consulta determinan si la vista materializada se puede actualizar de forma incremental o completa. Para obtener

información sobre el método de actualización, consulte REFRESH MATERIALIZED VIEW. Para obtener información acerca de las limitaciones para la actualización incremental, consulte Limitaciones para la actualización incremental.

Si la consulta contiene un comando SQL que no es compatible con la actualización progresiva, Amazon Redshift muestra un mensaje que indica que la vista materializada utilizará una actualización completa. El mensaje se puede mostrar o no, en función de la aplicación cliente de SQL. Active la columna `state` de STV_MV_INFO para ver el tipo de actualización utilizado por una vista materializada.

AUTO REFRESH

Se trata de una cláusula que define si la vista materializada debe actualizarse de forma automática con los últimos cambios de las tablas base. El valor predeterminado es `NO`. Para obtener más información, consulte Actualización de una vista materializada.

## Notas de uso

Para crear una vista materializada, debe tener los siguientes privilegios:

- Privilegios CREATE para un esquema.
- Privilegio SELECT de nivel de tabla o de columna en las tablas base para crear una vista materializada. Si tiene privilegios de nivel de columna en columnas específicas, no puede crear una vista materializada solo en esas columnas.

## Actualización incremental para obtener vistas materializadas en un recurso compartido de datos

Amazon Redshift admite la actualización automática e incremental de las vistas materializadas en un recurso compartido de datos de consumidores cuando se comparten las tablas base. La actualización incremental es una operación en la que Amazon Redshift identifica los cambios en la tabla o tablas base que se produjeron después de la actualización anterior y actualiza solo los registros correspondientes de la vista materializada. Esto se ejecuta más rápido que una actualización completa y mejora el rendimiento de la carga de trabajo. No es necesario cambiar la definición de la vista materializada para aprovechar la actualización incremental.

Hay un par de limitaciones que hay que tener en cuenta a la hora de aprovechar la actualización incremental con una vista materializada:

- La vista materializada debe hacer referencia solo a una base de datos, ya sea local o remota.
- La actualización incremental solo está disponible en las nuevas vistas materializadas. Por lo tanto, debe eliminar las vistas materializadas existentes y volver a crearlas para que se produzca una actualización incremental.

Para obtener más información sobre la creación de vistas materializadas en un recurso compartido de datos, consulte Trabajo con vistas en el uso compartido de datos de Amazon Redshift, que contiene varios ejemplos de consultas.

## Actualizaciones DDL en vistas materializadas o tablas base

Cuando utilice vistas materializadas en Amazon Redshift, siga las siguientes notas de uso para las actualizaciones de lenguaje de definición de datos (DDL) en vistas materializadas o tablas base.

- Puede agregar columnas a una tabla base sin que esto afecte a las vistas materializadas que hacen referencia a la tabla base.
- Algunas operaciones pueden dejar la vista materializada en un estado en el que no se puede actualizar. Por ejemplo, operaciones como renombrar o eliminar una columna, cambiar el tipo de columna y cambiar el nombre de un esquema. Dichas vistas materializadas se pueden consultar pero no se pueden actualizar. En este caso, debe eliminar y volver a crear la vista materializada.
- En general, no se puede modificar la definición de una vista materializada (su instrucción SQL).
- No se puede cambiar el nombre de una vista materializada.

## Limitaciones

No puede definir una vista materializada que haga referencia a o incluya cualquiera de los siguientes:

- Vistas estándar o vistas y tablas de sistema.
- Tablas temporales.
- Funciones definidas por el usuario.
- La cláusula ORDER BY, LIMIT u OFFSET.
- Referencias de enlace tardío a tablas básicas. En otras palabras, cualquier tabla de base o columnas relacionadas a las que se hace referencia en la definición de la consulta SQL de la vista materializada debe existir y ser válida.
- Funciones específicas del nodo principal: CURRENT_SCHEMA, CURRENT_SCHEMAS, HAS_DATABASE_PRIVILEGE, HAS_SCHEMA_PRIVILEGE, HAS_TABLE_PRIVILEGE.

- No se puede utilizar la opción AUTO REFRESH YES cuando la definición de la vista materializada incluye funciones mutables o esquemas externos. Tampoco puede usarla cuando define una vista materializada en otra vista materializada.
- No hace falta que ejecute manualmente ANALYZE en las vistas materializadas. Actualmente, esto solo ocurre mediante AUTO ANALYZE. Para obtener más información, consulte Análisis de tablas.
- Tablas protegidas por RLS o DDM.

## Ejemplos

En el siguiente ejemplo, se crea una vista materializada a partir de tres tablas base que se combinan y se agrupan. Cada fila representa una categoría con el número de entradas vendidas. Cuando consulta la vista materializada tickets_mv, accede directamente a los datos precalculados en la vista materializada tickets_mv.

```
CREATE MATERIALIZED VIEW tickets_mv AS
    select   catgroup,
    sum(qtysold) as sold
    from     category c, event e, sales s
    where    c.catid = e.catid
    and      e.eventid = s.eventid
    group by catgroup;
```

En el siguiente ejemplo, se crea una vista materializada similar al ejemplo anterior y se utiliza la función agregada MAX().

```
CREATE MATERIALIZED VIEW tickets_mv_max AS
    select   catgroup,
    max(qtysold) as sold
    from     category c, event e, sales s
    where    c.catid = e.catid
    and      e.eventid = s.eventid
    group by catgroup;

SELECT name, state FROM STV_MV_INFO;
```

En el siguiente ejemplo, se utiliza una cláusula UNION ALL para combinar la tabla `public_sales` de Amazon Redshift y la tabla `spectrum.sales` de Redshift Spectrum para crear una vista material `mv_sales_vw`. Para obtener información acerca del comando CREATE EXTERNAL

TABLE para Amazon Redshift Spectrum, consulte CREATE EXTERNAL TABLE. La tabla externa de Redshift Spectrum referencia los datos de Amazon S3.

```
CREATE MATERIALIZED VIEW mv_sales_vw as
select salesid, qtysold, pricepaid, commission, saletime from public.sales
union all
select salesid, qtysold, pricepaid, commission, saletime from spectrum.sales
```

En el siguiente ejemplo, se crea una vista materializada `mv_fq` basada en una tabla externa de consulta federada. Para obtener información acerca de las consultas federadas, consulte CREATE EXTERNAL SCHEMA.

```
CREATE MATERIALIZED VIEW mv_fq as select firstname, lastname from apg.mv_fq_example;

select firstname, lastname from mv_fq;
 firstname | lastname
-----------+----------
 John      | Day
 Jane      | Doe
(2 rows)
```

En el siguiente ejemplo, se muestra la definición de una vista materializada.

```
SELECT pg_catalog.pg_get_viewdef('mv_sales_vw'::regclass::oid, true);

pg_get_viewdef
------------------------------------------------
create materialized view mv_sales_vw as select a from t;
```

El siguiente ejemplo muestra cómo configurar AUTO REFRESH en la definición de vista materializada y también especifica un DISTSTYLE. Primero, crea una tabla base simple.

```
CREATE TABLE baseball_table (ball int, bat int);
```

A continuación, cree una vista materializada.

```
CREATE MATERIALIZED VIEW mv_baseball DISTSTYLE ALL AUTO REFRESH YES AS SELECT ball AS
 baseball FROM baseball_table;
```

Ahora puede consultar la vista materializada de mv_baseball. Para comprobar si la actualización automática está activada en una vista materializada, consulte STV_MV_INFO.

En el siguiente ejemplo se crea una vista materializada que hace referencia a una tabla de origen de otra base de datos. Se supone que la base de datos que contiene la tabla de origen, database_A, se encuentra en el mismo clúster o grupo de trabajo que la vista materializada, que se crea en database_B. (Puede sustituir el ejemplo por sus propias bases de datos). Primero, cree una tabla en database_A denominada cities, con una columna de cityname. Haga que el tipo de datos de la columna sea VARCHAR. Después de crear la tabla de origen, ejecute el siguiente comando en database_B para crear una vista materializada cuyo origen sea la tabla cities. Asegúrese de especificar la base de datos y el esquema de la tabla de origen en la cláusula FROM:

```
CREATE MATERIALIZED VIEW cities_mv AS
SELECT  cityname
FROM    database_A.public.cities;
```

Consulte la vista materializada que ha creado. La consulta recupera los registros cuyo origen original es la tabla cities de database_A:

```
select * from cities_mv;
```

Al ejecutar la instrucción SELECT, cities_mv devuelve los registros. Los registros se actualizan desde la tabla de origen solo cuando se ejecuta una instrucción REFRESH. Además, tenga en cuenta que no puede actualizar los registros directamente en la vista materializada. Para obtener información sobre cómo actualizar los datos en una vista materializada, consulte REFRESH MATERIALIZED VIEW.

Para obtener información detallada sobre la información general de la vista materializada y los comandos SQL utilizados para actualizar y eliminar vistas materializadas, consulte los siguientes temas:

- Vistas materializadas en Amazon Redshift
- REFRESH MATERIALIZED VIEW
- DROP MATERIALIZED VIEW

## CREATE MODEL

Temas

- Requisitos previos
- Privilegios necesarios
- Control de costes
- CREATE MODEL completo
- Parámetros
- Notas de uso
- Casos de uso


## Requisitos previos

Antes de utilizar la instrucción CREATE MODEL, complete los requisitos previos indicados en Configuración del clúster para utilizar Amazon Redshift ML. A continuación, se brinda un resumen de alto nivel de los requisitos previos.

- Cree un clúster de Amazon Redshift con la consola de administración de AWS o la AWS Command Line Interface (AWS CLI)
- Adjunte la política de AWS Identity and Access Management (IAM) mientras crea el clúster.
- Para permitir que Amazon Redshift y SageMaker AI asuman el rol para interactuar con otros servicios, añada la política de confianza adecuada al rol de IAM.

Para obtener detalles sobre el rol de IAM, la política de confianza y otros requisitos previos, consulte Configuración del clúster para utilizar Amazon Redshift ML.

A continuación, encontrará diferentes casos de uso para la instrucción CREATE MODEL.


- CREATE MODEL simple
- CREATE MODEL con guía para el usuario
- CREATE para modelos XGBoost con AUTO OFF
- Bring your own model (BYOM): inferencia local
- Bring your own model (BYOM): inferencia remota
- CREATE MODEL con K-MEANS
- CREATE MODEL completo

## Privilegios necesarios

Los siguientes privilegios son necesarios para CREATE MODEL:

- Superusuario
- Usuarios con el privilegio CREATE MODEL
- Roles con el privilegio GRANT CREATE MODEL

## Control de costes

Amazon Redshift ML utiliza los recursos de clúster existentes para crear modelos de predicción, por lo que no es necesario pagar costes adicionales. Sin embargo, es posible que tenga costes adicionales si necesita cambiar el tamaño del clúster o desea entrenar los modelos. Amazon Redshift ML utiliza Amazon SageMaker AI para entrenar modelos, lo que supone un costo adicional asociado. Hay formas de controlar los costes adicionales, como limitar el tiempo máximo que puede tardar el entrenamiento o limitar el número de ejemplos de formación utilizados a la hora de entrenar el modelo. Para obtener más información, consulte Costes para usar Amazon Redshift ML.

## CREATE MODEL completo

A continuación, se resumen las opciones básicas de la sintaxis de CREATE MODEL completo.

Sintaxis de CREATE MODEL completo

A continuación, se muestra la sintaxis completa de la instrucción CREATE MODEL.

> **Important**
>
> Cuando cree un modelo con la instrucción CREATE MODEL, siga el orden de las palabras clave en la sintaxis que aparece a continuación.

```
CREATE MODEL model_name
FROM { table_name | ( select_statement )  | 'job_name' }
[ TARGET column_name ]
FUNCTION function_name [ ( data_type [, ...] ) ]
[ RETURNS data_type ]
  -- supported only for BYOM
[ SAGEMAKER 'endpoint_name'[:'model_name']]
  -- supported only for BYOM remote inference
```

```
IAM_ROLE { default | 'arn:aws:iam::<account-id>:role/<role-name>' }
[ AUTO ON / OFF ]
  -- default is AUTO ON
[ MODEL_TYPE { XGBOOST | MLP | LINEAR_LEARNER | KMEANS | FORECAST } ]
  -- not required for non AUTO OFF case, default is the list of all supported types
  -- required for AUTO OFF
[ PROBLEM_TYPE ( REGRESSION | BINARY_CLASSIFICATION | MULTICLASS_CLASSIFICATION ) ]
  -- not supported when AUTO OFF
[ OBJECTIVE ( 'MSE' | 'Accuracy' | 'F1' | 'F1_Macro' | 'AUC' |
              'reg:squarederror' | 'reg:squaredlogerror'| 'reg:logistic'|
              'reg:pseudohubererror' | 'reg:tweedie' | 'binary:logistic' |
 'binary:hinge',
              'multi:softmax' | 'RMSE' | 'WAPE' | 'MAPE' | 'MASE' |
 'AverageWeightedQuantileLoss' ) ]
  -- for AUTO ON: first 5 are valid
  -- for AUTO OFF: 6-13 are valid
  -- for FORECAST: 14-18 are valid
[ PREPROCESSORS 'string' ]
  -- required for AUTO OFF, when it has to be 'none'
  -- optional for AUTO ON
[ HYPERPARAMETERS { DEFAULT | DEFAULT EXCEPT ( Key 'value' (,...) ) } ]
  -- support XGBoost hyperparameters, except OBJECTIVE
  -- required and only allowed for AUTO OFF
  -- default NUM_ROUND is 100
  -- NUM_CLASS is required if objective is multi:softmax (only possible for AUTO OFF)
 [ SETTINGS (
   S3_BUCKET 'amzn-s3-demo-bucket',  |
     -- required
  TAGS 'string', |
     -- optional
  KMS_KEY_ID 'kms_string', |
     -- optional
  S3_GARBAGE_COLLECT on / off, |
     -- optional, defualt is on.
  MAX_CELLS integer, |
     -- optional, default is 1,000,000
  MAX_RUNTIME integer (, ...) |
     -- optional, default is 5400 (1.5 hours)
  HORIZON integer, |
     -- required if creating a forecast model
  FREQUENCY integer, |
     -- required if creating a forecast model
  PERCENTILES string, |
     -- optional if creating a forecast model
```

```
  MAX_BATCH_ROWS integer -- optional for BYOM remote inference
    ) ]
```

## Parámetros

model_name

El nombre del modelo. El nombre del modelo en un esquema debe ser único.

FROM { table_name | ( select_query ) | 'job_name'}

Se trata del nombre de la tabla o la consulta que especifica los datos de formación. Pueden ser una tabla existente en el sistema o una consulta SELECT compatible con Amazon Redshift entre paréntesis, es decir, (). En el resultado de la consulta, debe haber al menos dos columnas.

TARGET column_name

Se trata del nombre de la columna que se convierte en el objetivo de predicción. La columna debe existir en la cláusula FROM.

FUNCTION function_name ( data_type [, ...] )

Se trata del nombre de la función que se creará y los tipos de datos de los argumentos de entrada. Puede proporcionar el nombre de un esquema de la base de datos en lugar del nombre de una función.

RETURNS data_type

El tipo de datos que se va a devolver desde la función del modelo. El tipo de datos SUPER de devolución solo se aplica a BYOM con inferencia remota.

SAGEMAKER 'endpoint_name'[:'model_name']

El nombre del punto de conexión de Amazon SageMaker AI. Si el nombre del punto de conexión apunta a un punto de conexión de múltiples modelos, agregue el nombre del modelo que se utilizará. El punto de conexión debe estar alojado en la misma Región de AWS que el clúster de Amazon Redshift.

IAM_ROLE { default | 'arn:aws:iam::<account-id>:role/<role-name>'}

Utilice la palabra clave predeterminada para que Amazon Redshift utilice el rol de IAM configurado como predeterminado y asociado al clúster cuando se ejecuta el comando CREATE MODEL. También puede especificar un ARN de un rol de IAM para usar ese rol.

[AUTO ON/OFF]

Activa o desactiva la detección automática de CREATE MODEL de procesadores previos, algoritmos e hiperparámetros. Si especifica esta opción al crear un modelo de previsión, se debe utilizar un AutoPredictor, en el que Amazon Forecast aplica las combinaciones óptimas de algoritmos a cada serie temporal del conjunto de datos.

MODEL_TYPE { XGBOOST | MLP | LINEAR_LEARNER | KMEANS | FORECAST }

(Opcional) Especifica el tipo de modelo. Puede especificar si desea entrenar un modelo de un tipo específico, como XGBoost, perceptron multicapa (MLP), KMEANS o Linear Learner, que son todos algoritmos compatibles con el Piloto automático de Amazon SageMaker AI. Si no se especifica el parámetro, se buscan todos los tipos de modelos admitidos durante la formación para encontrar el mejor modelo. También puede crear un modelo de previsión en Redshift ML para crear previsiones precisas de series temporales.

PROBLEM_TYPE ( REGRESSION | BINARY_CLASSIFICATION | MULTICLASS_CLASSIFICATION )

(Opcional) Especifica el tipo de problema. Si conoce el tipo de problema, puede restringir a Amazon Redshift a que busque solo el mejor modelo de ese tipo específico. Si no se especifica este parámetro, se detecta un tipo de problema durante la formación, en función de sus datos.

OBJECTIVE ( 'MSE' | 'Accuracy' | 'F1' | 'F1Macro' | 'AUC' | 'reg:squarederror' | 'reg:squaredlogerror' | 'reg:logistic' | 'reg:pseudohubererror' | 'reg:tweedie' | 'binary:logistic' | 'binary:hinge' | 'multi:softmax' | 'RMSE' | 'WAPE' | 'MAPE' | 'MASE' | 'AverageWeightedQuantileLoss' )

(Opcional) Especifica el nombre de la métrica de objetivo utilizada para medir la calidad predictiva de un sistema de machine learning. Esta métrica se optimiza durante la formación para proporcionar la mejor estimación para los valores del parámetro del modelo a partir de los datos. Si no se especifica una métrica de forma explícita, el comportamiento predeterminado es utilizar automáticamente MSE para regresión, F1 para clasificación binaria y Accuracy para clasificación multiclase. Para obtener más información acerca de los objetivos, consulte AutoMLJobObjective en la Referencia de la API de Amazon SageMaker AI y Learning task parameters en la documentación de XGBOOST. Los valores RMSE, WAPE, MAPE, MASE y AverageWeightedQuantileLoss solo son aplicables a los modelos de previsión. Para obtener más información, consulte la operación de la API CreateAutoPredictor.

PREPROCESSORS "string"

(Opcional) Especifica ciertas combinaciones de procesadores previos para determinados conjuntos de columnas. El formato es una lista de columnSets y las transformaciones adecuadas

que se aplicarán a cada conjunto de columnas. Amazon Redshift aplica todos los transformadores de una lista de transformadores específica a todas las columnas del ColumnSet correspondiente. Por ejemplo, para aplicar OneHotEncoder con Imputer a las columnas t1 y t2, utilice el comando de muestra que aparece a continuación.

```
CREATE MODEL customer_churn
FROM customer_data
TARGET 'Churn'
FUNCTION predict_churn
IAM_ROLE { default | 'arn:aws:iam::<account-id>:role/<role-name>' }
PROBLEM_TYPE BINARY_CLASSIFICATION
OBJECTIVE 'F1'
PREPROCESSORS '[
...
  {"ColumnSet": [
      "t1",
      "t2"
    ],
    "Transformers": [
      "OneHotEncoder",
      "Imputer"
    ]
  },
  {"ColumnSet": [
      "t3"
    ],
    "Transformers": [
      "OneHotEncoder"
    ]
  },
  {"ColumnSet": [
      "temp"
    ],
    "Transformers": [
      "Imputer",
      "NumericPassthrough"
    ]
  }
]'
SETTINGS (
  S3_BUCKET 'amzn-s3-demo-bucket'
)
```

HYPERPARAMETERS { DEFAULT | DEFAULT EXCEPT ( key ‘value’ (,..) ) }

Especifica si los parámetros de XGBoost predeterminados se utilizan o se sustituyen con valores especificados por el usuario. Los valores deben ir entre comillas simples. Los siguientes son ejemplos de parámetros para XGBoost y sus valores predeterminados.

| Nombre del parámetro | Valor del parámetro | Valor predeterminado | Notas |
|---|---|---|---|
| num_class | Entero | Es obligatorio para la clasificación multiclase. | N/A |
| num_round | Entero | 100 | N/A |
| tree_method | Cadena | Automático | N/A |
| max_depth | Entero | 6 | [0, 10] |
| min_child_weight | Flotante | 1 | MinValue: 0, MaxValue: 120 |
| subsample | Flotante | 1 | MinValue: 0,5, MaxValue: 1 |
| gamma | Flotante | 0 | MinValue: 0, MaxValue: 5 |
| alpha | Flotante | 0 | MinValue: 0, MaxValue: 1000 |
| eta | Flotante | 0.3 | MinValue: 0,1, MaxValue: 0,5 |

| Nombre del parámetro | Valor del parámetro | Valor predeterminado | Notas |
|---|---|---|---|
| colsample_byleve | Flotante | 1 | MinValue: 0,1, MaxValue: 1 |
| colsample_bynode | Flotante | 1 | MinValue: 0,1, MaxValue: 1 |
| colsample_bytree | Flotante | 1 | MinValue: 0,5, MaxValue: 1 |
| lambda | Flotante | 1 | MinValue: 0, MaxValue: 1000 |
| max_delta_step | Entero | 0 | [0, 10] |

SETTINGS ( S3_BUCKET 'amzn-s3-demo-bucket', | TAGS 'string', | KMS_KEY_ID 'kms_string' , | S3_GARBAGE_COLLECT on / off, | MAX_CELLS integer , | MAX_RUNTIME (,...) , | HORIZON integer, | FREQUENCY forecast_frequency, | PERCENTILES matriz de cadenas )

La cláusula S3_BUCKET especifica la ubicación de Amazon S3 que se utiliza para almacenar resultados intermedios.

(Opcional) El parámetro TAGS es una lista separada por comas de pares clave-valor que puede utilizar para etiquetar recursos creados en Amazon SageMaker AI y Amazon Forecast. Las etiquetas lo ayudan a organizar los recursos y a asignar los costos. Los valores del par son opcionales, por lo que puede crear etiquetas mediante el formato `key=value` o simplemente con la creación de una clave. Para obtener más información sobre las etiquetas en Amazon Redshift, consulte Información general del proceso de etiquetado.

(Opcional) KMS_KEY_ID especifica si Amazon Redshift utiliza cifrado del lado del servidor con una clave de AWS KMS para proteger los datos en reposo. Los datos en tránsito están protegidos con la capa de conexión segura (SSL).

(Opcional) S3_GARBAGE_COLLECT {ON|OFF} especifica si Amazon Redshift realiza la recopilación de elementos no utilizados en los conjuntos de datos resultantes utilizados para formar modelos, además de los propios modelos. Si se establece en OFF, los conjuntos de

datos resultantes utilizados para formar modelos y los modelos permanecen en Amazon S3 y se pueden utilizar para otros fines. Si se establece en ON, Amazon Redshift elimina los artefactos en Amazon S3 una vez que se completa la formación. El valor predeterminado es ON.

(Opcional) MAX_CELLS especifica el número de celdas de los datos de formación. Este valor se obtiene de multiplicar el número de registros (en la tabla o en la consulta de formación) por el número de columnas. El valor predeterminado es 1 000 000.

(Opcional) MAX_RUNTIME especifica la cantidad máxima de tiempo que dura la formación. Los trabajos de formación suelen completarse antes en función del tamaño del conjunto de datos. Especifica la cantidad máxima de tiempo que debería tomar la formación. El valor predeterminado es 5400 (90 minutos).

HORIZON especifica el número máximo de predicciones que puede devolver el modelo de previsión. Una vez que el modelo esté formado, no podrá cambiar este entero. Este parámetro es obligatorio si se entrena un modelo de previsión.

FREQUENCY especifica los detalles en unidades de tiempo que desea que sean las previsiones. Las opciones disponibles son `Y | M | W | D | H | 30min | 15min | 10min | 5min | 1min`. Este parámetro es obligatorio si se entrena un modelo de previsión.

(Opcional) PERCENTILES es una cadena delimitada por comas que especifica los tipos de previsión que se utilizan para entrenar un predictor. Los tipos de previsión pueden ser cuantiles de 0,01 a 0,99, en incrementos de 0,01 o más. También puede especificar la previsión media con la media. Puede especificar un máximo de cinco tipos de previsión.

MAX_BATCH_ROWS integer

(Opcional) El número máximo de filas que Amazon Redshift envía en una sola solicitud por lotes para una única invocación de SageMaker AI. Solo se admite para BYOM con inferencia remota. El valor mínimo de este parámetro es 1. El valor máximo es `INT_MAX` o 2 147 483 647. Este parámetro solo es necesario cuando el tipo de datos de entrada y de devolución son SUPER. El valor predeterminado es `INT_MAX` o 2 147 483 647.

## Notas de uso

Cuando utilice CREATE MODEL, tenga en cuenta lo siguiente:

- La instrucción CREATE MODEL funciona en modo asíncrono y devuelve los resultados una vez exportados los datos de formación a Amazon S3. Los pasos restantes de entrenamiento en

Amazon SageMaker AI se producen en segundo plano. Mientras la formación está en curso, la función de inferencia correspondiente está visible, pero no se puede ejecutar. Puede consultar STV_ML_MODEL_INFO para ver el estado de la formación.

- La formación puede durar hasta 90 minutos en segundo plano de manera predeterminada en el modelo Auto y puede extenderse. Para cancelar la formación, simplemente ejecute el comando DROP MODEL.
- El clúster de Amazon Redshift que se utiliza a la hora de crear el modelo y el bucket de Amazon S3 que se utiliza para preparar los datos de formación y los artefactos del modelo deberán estar en la misma región de AWS.
- Durante el entrenamiento del modelo, Amazon Redshift y SageMaker AI almacenan artefactos intermedios en el bucket de Amazon S3 proporcionado. De manera predeterminada, Amazon Redshift realiza la recopilación de elementos no utilizados al final de la operación CREATE MODEL. Amazon Redshift quita esos objetos de Amazon S3. Para retener esos artefactos en Amazon S3, establezca la opción S3_GARBAGE COLLECT OFF.
- Debe utilizar al menos 500 filas de los datos de formación proporcionados en la cláusula FROM.
- Solo se pueden especificar hasta 256 columnas de características (entrada) en la cláusula FROM { table_name | ( select_query ) } cuando se utiliza la instrucción CREATE MODEL.
- Para AUTO ON, los tipos de columnas que puede utilizar como conjunto de formación son SMALLINT, INTEGER, BIGINT, DECIMAL, REAL, DOUBLE, BOOLEAN, CHAR, VARCHAR, DATE, TIME, TIMETZ, TIMESTAMP y TIMESTAMPTZ. Para AUTO OFF, los tipos de columnas que puede utilizar como conjunto de formación son SMALLINT, INTEGER, BIGINT, DECIMAL, REAL, DOUBLE y BOOLEAN.
- No se puede utilizar DECIMAL, DATE, TIME, TIMETZ, TIMESTAMP, TIMESTAMPTZ, GEOMETRY, GEOGRAPHY, HLLSKETCH, SUPER, o VARBYTE como el tipo de columna de destino.
- Para mejorar la precisión del modelo, lleve a cabo una las siguientes acciones:
  - Cuando especifique los datos de formación en la cláusula FROM, agregue tantas columnas relevantes en el comando CREATE MODEL como sea posible.
  - Utilice un valor más grande para MAX_RUNTIME y MAX_CELLS. Los valores más grandes para este parámetro aumentan el costo de la formación de un modelo.
- La ejecución de la instrucción CREATE MODEL presenta los resultados tan pronto como los datos de formación se calculan y exportan al bucket de Amazon S3. Después de ese punto, se puede verificar el estado de la formación mediante el comando SHOW MODEL. Cuando un modelo que se está formando en segundo plano presenta error, se puede verificar el error con SHOW MODEL.

No se puede volver a probar un modelo que presente error. Utilice DROP MODEL para quitar un modelo que presenta error y volver a crear un modelo nuevo. Para obtener más información acerca de SHOW MODEL, consulte SHOW MODEL.

- El procedimiento BYOM local es compatible con el mismo tipo de modelos que Amazon Redshift ML admite para casos que no son de BYOM. Amazon Redshift es compatible con modelos de XGBoost (mediante la versión 1.0 o posteriores de XGBoost) y KMEANS básicos sin preprocesadores y modelos de XGBOOST/MLP/Linear Learner entrenados por el Piloto automático de Amazon SageMaker AI. Admite estos últimos modelos con procesadores previos que el Piloto automático ha especificado y que también son compatibles con Amazon SageMaker AI Neo.
- Si su clúster de Amazon Redshift ha habilitado el enrutamiento mejorado para su nube privada virtual (VPC), asegúrese de crear un punto de conexión de VPC de Amazon S3 y un punto de conexión de VPC de SageMaker AI para la VPC en la que se encuentra el clúster. Esto permite que el tráfico pase por la VPC entre estos servicios durante el proceso CREATE MODEL. Para obtener más información, consulte SageMaker AI Clarify Job Amazon VPC Subnets and Security Groups.

## Casos de uso

En los siguientes casos de uso se muestra cómo utilizar CREATE MODEL a la medida de sus necesidades.

### CREATE MODEL simple

A continuación, se resumen las opciones básicas de la sintaxis de CREATE MODEL.

Sintaxis de CREATE MODEL simple

```
CREATE MODEL model_name
FROM { table_name | ( select_query ) }
TARGET column_name
FUNCTION prediction_function_name
IAM_ROLE { default }
SETTINGS (
  S3_BUCKET 'amzn-s3-demo-bucket',
  [ MAX_CELLS integer ]
)
```

Parámetros de CREATE MODEL simple

model_name

El nombre del modelo. El nombre del modelo en un esquema debe ser único.

FROM { table_name | ( select_query ) }

Se trata del nombre de la tabla o la consulta que especifica los datos de formación. Pueden ser una tabla existente en el sistema o una consulta SELECT compatible con Amazon Redshift entre paréntesis, es decir, (). En el resultado de la consulta, debe haber al menos dos columnas.

TARGET column_name

Se trata del nombre de la columna que se convierte en el objetivo de predicción. La columna debe existir en la cláusula FROM.

FUNCTION prediction_function_name

Se trata de un valor que especifica el nombre de la función de machine learning de Amazon Redshift que generará CREATE MODEL y que se utilizará para generar predicciones con este modelo. La función se crea en el mismo esquema que el objeto del modelo y se puede sobrecargar.

El machine learning de Amazon Redshift es compatible con modelos, como el árbol Xtreme Gradient Boosted (XGBoost) para la regresión y la clasificación.

IAM_ROLE { default | 'arn:aws:iam::<account-id>:role/<role-name>' }

Utilice la palabra clave predeterminada para que Amazon Redshift utilice el rol de IAM configurado como predeterminado y asociado al clúster cuando se ejecuta el comando CREAT MODEL. También puede especificar el ARN de un rol de IAM para usar ese rol.

S3_BUCKET 'amzn-s3-demo-bucket'

Se trata del nombre del bucket de Amazon S3 que creó anteriormente y que se utiliza para compartir datos de entrenamiento y artefactos entre Amazon Redshift y SageMaker AI. Amazon Redshift crea una subcarpeta en este bucket antes de descargar los datos de formación. Una vez que se completa la formación, Amazon Redshift elimina la subcarpeta creada y su contenido.

Número entero para MAX_CELLS

Se trata del número máximo de celdas que se exportarán desde la cláusula FROM. El valor predeterminado es 1 000 000.

El número de celdas se obtiene de multiplicar el número de filas de los datos de formación (productos de la tabla o la consulta de la cláusula FROM) por el número de columnas. Si el número de celdas de los datos de formación es mayor que el especificado por el parámetro max_cells, CREATE MODEL reduce la cantidad de datos de formación de la cláusula FROM para reducir el tamaño del conjunto de formación debajo de MAX_CELLS. Si se admiten conjuntos de datos de formación más grandes pueden brindar mayor precisión, pero también puede significar que el modelo tarde más en formarse y cueste más.

Para obtener información acerca de los costos que implica usar Amazon Redshift, consulte Costos de utilizar Amazon Redshift ML.

Para obtener más información sobre los costos asociados a los distintos números de celdas y los detalles de la prueba gratuita, consulte Precios de Amazon Redshift.

CREATE MODEL con guía para el usuario

A continuación, encontrará una descripción de las opciones de CREATE MODEL, además de las opciones descritas en CREATE MODEL simple.

De manera predeterminada, CREATE MODEL busca la mejor combinación de procesamiento previo y modelo para su conjunto de datos específico. Es posible que desee un control adicional o introducir conocimientos de dominio adicionales (como el tipo de problema o el objetivo) sobre su modelo. En un escenario de deserción de clientes, si el resultado “el cliente no está activo” es poco frecuente, se suele preferir el objetivo F1 en lugar del objetivo de precisión. Como los modelos de alta precisión tal vez predicen que “el cliente está activo” en todo momento, esto se traduce en una alta precisión, pero escaso valor empresarial. Para obtener información acerca del objetivo F1, consulte AutoMLJobObjective en la Referencia de la API de Amazon SageMaker AI.

Luego, CREATE MODEL sigue las sugerencias sobre los aspectos especificados, como el objetivo. Al mismo tiempo, CREATE MODEL descubre de forma automática los mejores procesadores previos y los mejores hiperparámetros.

CREATE MODEL con sintaxis de guía para el usuario

CREATE MODEL ofrece más flexibilidad en los aspectos que se pueden especificar y en los aspectos que Amazon Redshift descubre de forma automática.

```
CREATE MODEL model_name
FROM { table_name | ( select_statement ) }
```

```
TARGET column_name
FUNCTION function_name
IAM_ROLE { default }
[ MODEL_TYPE { XGBOOST | MLP | LINEAR_LEARNER} ]
[ PROBLEM_TYPE ( REGRESSION | BINARY_CLASSIFICATION | MULTICLASS_CLASSIFICATION ) ]
[ OBJECTIVE ( 'MSE' | 'Accuracy' | 'F1' | 'F1Macro' | 'AUC') ]
SETTINGS (
  S3_BUCKET 'amzn-s3-demo-bucket', |
  S3_GARBAGE_COLLECT { ON | OFF }, |
  KMS_KEY_ID 'kms_key_id', |
  MAX_CELLS integer, |
  MAX_RUNTIME integer (, ...)
)
```

CREATE MODEL con parámetros de guía para el usuario

MODEL_TYPE { XGBOOST | MLP | LINEAR_LEARNER }

(Opcional) Especifica el tipo de modelo. Puede especificar si desea entrenar un modelo de un tipo específico, como XGBoost, perceptron multicapa (MLP) o Linear Learner, que son todos algoritmos compatibles con el Piloto automático de Amazon SageMaker AI. Si no se especifica el parámetro, se buscan todos los tipos de modelos admitidos durante la formación para encontrar el mejor modelo.

PROBLEM_TYPE ( REGRESSION | BINARY_CLASSIFICATION | MULTICLASS_CLASSIFICATION )

(Opcional) Especifica el tipo de problema. Si conoce el tipo de problema, puede restringir a Amazon Redshift a que busque solo el mejor modelo de ese tipo específico. Si no se especifica este parámetro, se detecta un tipo de problema durante la formación, en función de sus datos.

OBJECTIVE ( 'MSE' | 'Accuracy' | 'F1' | 'F1Macro' | 'AUC')

(Opcional) Especifica el nombre de la métrica de objetivo utilizada para medir la calidad predictiva de un sistema de machine learning. Esta métrica se optimiza durante la formación para proporcionar la mejor estimación para los valores del parámetro del modelo a partir de los datos. Si no se especifica una métrica de forma explícita, el comportamiento predeterminado es utilizar automáticamente MSE para regresión, F1 para clasificación binaria y Accuracy para clasificación multiclase. Para obtener más información acerca de los objetivos, consulte AutoMLJobObjective en la Referencia de la API de Amazon SageMaker AI.

Número entero para MAX_CELLS

(Opcional) Especifica el número de celdas de los datos de formación. Este valor se obtiene de multiplicar el número de registros (en la tabla o en la consulta de formación) por el número de columnas. El valor predeterminado es 1 000 000.

Número entero para MAX_RUNTIME

(Opcional) Especifica la cantidad máxima de tiempo que dura la formación. Los trabajos de formación suelen completarse antes en función del tamaño del conjunto de datos. Especifica la cantidad máxima de tiempo que debería tomar la formación. El valor predeterminado es 5400 (90 minutos).

S3_GARBAGE_COLLECT { ON | OFF }

(Opcional) Especifica si Amazon Redshift realiza la recolección de basura en los conjuntos de datos resultantes utilizados para formar modelos, además de los propios modelos. Si se establece en OFF, los conjuntos de datos resultantes utilizados para formar modelos y los modelos permanecen en Amazon S3 y se pueden utilizar para otros fines. Si se establece en ON, Amazon Redshift elimina los artefactos en Amazon S3 una vez que se completa la formación. El valor predeterminado es ON.

KMS_KEY_ID 'kms_key_id'

(Opcional) Especifica si Amazon Redshift utiliza el cifrado del lado del servidor con una clave de AWS KMS para proteger los datos en reposo. Los datos en tránsito están protegidos con la capa de conexión segura (SSL).

PREPROCESSORS "string"

(Opcional) Especifica ciertas combinaciones de procesadores previos para determinados conjuntos de columnas. El formato es una lista de columnSets y las transformaciones adecuadas que se aplicarán a cada conjunto de columnas. Amazon Redshift aplica todos los transformadores de una lista de transformadores específica a todas las columnas del ColumnSet correspondiente. Por ejemplo, para aplicar OneHotEncoder con Imputer a las columnas t1 y t2, utilice el comando de muestra que aparece a continuación.

```
CREATE MODEL customer_churn
FROM customer_data
TARGET 'Churn'
FUNCTION predict_churn
IAM_ROLE { default | 'arn:aws:iam::<account-id>:role/<role-name>' }
PROBLEM_TYPE BINARY_CLASSIFICATION
```

```
OBJECTIVE 'F1'
PREPROCESSORS '[
...
{"ColumnSet": [
    "t1",
    "t2"
  ],
  "Transformers": [
    "OneHotEncoder",
    "Imputer"
  ]
},
{"ColumnSet": [
    "t3"
  ],
  "Transformers": [
    "OneHotEncoder"
  ]
},
{"ColumnSet": [
    "temp"
  ],
  "Transformers": [
    "Imputer",
    "NumericPassthrough"
  ]
}
]'
SETTINGS (
S3_BUCKET 'amzn-s3-demo-bucket'
)
```

Amazon Redshift es compatible con los siguientes transformadores:

- OneHotEncoder: generalmente se utiliza para codificar un valor discreto en un vector binario con un valor distinto de cero. Este transformador es adecuado para muchos modelos de machine learning.
- OrdinalEncoder: codifica valores discretos en un solo valor entero. Este transformador es adecuado para determinados modelos de machine learning, como MLP y Linear Learner.
- NumericPassthrough: transmite la entrada como está al modelo.
- Imputer: completa los valores que faltan y los valores que no son un número (Not a Number, NaN).

- ImputerWithIndicator: completa los valores que faltan y los valores NaN. Este transformador también crea un indicador que señala si existen valores que faltaban o que se completaron.
- Normalizer: normaliza los valores, lo que puede mejorar el rendimiento de muchos algoritmos de machine learning.
- DateTimeVectorizer: crea una integración de vectores, que representa una columna de tipo de datos de fecha y hora que se puede utilizar en modelos de machine learning.
- PCA: proyecta los datos en un espacio de dimensiones más pequeñas para reducir el número de características y, al mismo tiempo, conservar la mayor cantidad de información posible.
- StandardScaler: estandariza las características mediante la eliminación de la media y el escalado a la varianza unitaria.
- MinMax: transforma las características mediante el escalado de cada característica a un rango determinado.

Amazon Redshift ML almacena los transformadores formados y los aplica de forma automática como parte de la consulta de predicción. No es necesario especificarlos cuando se generan predicciones a partir del modelo.

CREATE para modelos XGBoost con AUTO OFF

El comando AUTO OFF CREATE MODEL tiene objetivos generalmente diferentes a los del comando CREATE MODEL predeterminado.

Como usuario avanzado que ya conoce el tipo de modelo que desea y los hiperparámetros que se utilizan para formar estos modelos, puede utilizar CREATE MODEL con AUTO OFF para desactivar la detección automática de CREATE MODEL de procesadores previos e hiperparámetros. Para ello, especifica de forma explícita el tipo de modelo. Actualmente, XGBoost es el único tipo de modelo compatible cuando AUTO está configurado como OFF. Puede especificar hiperparámetros. Amazon Redshift utiliza valores predeterminados para cualquier hiperparámetro que especifique.

CREATE para modelos XGBoost con sintaxis AUTO OFF

```
CREATE MODEL model_name
FROM { table_name | (select_statement ) }
TARGET column_name
FUNCTION function_name
IAM_ROLE { default }
AUTO OFF
MODEL_TYPE XGBOOST
```

```
OBJECTIVE { 'reg:squarederror' | 'reg:squaredlogerror' | 'reg:logistic' |
            'reg:pseudohubererror' | 'reg:tweedie' | 'binary:logistic' | 'binary:hinge'
 |
            'multi:softmax' | 'rank:pairwise' | 'rank:ndcg' }
HYPERPARAMETERS DEFAULT EXCEPT (
    NUM_ROUND '10',
    ETA '0.2',
    NUM_CLASS '10',
    (, ...)
)
PREPROCESSORS 'none'
SETTINGS (
  S3_BUCKET 'amzn-s3-demo-bucket', |
  S3_GARBAGE_COLLECT { ON | OFF }, |
  KMS_KEY_ID 'kms_key_id', |
  MAX_CELLS integer, |
  MAX_RUNTIME integer (, ...)
)
```

CREATE para modelos XGBoost con parámetros AUTO OFF

AUTO OFF

Desactiva la detección automática de CREATE MODEL de procesadores previos, algoritmos e hiperparámetros.

MODEL_TYPE XGBOOST

Especifica el uso de XGBOOST para formar el modelo.

OBJECTIVE str

Especifica un objetivo reconocido por el algoritmo. Amazon Redshift admite reg:squarederror, reg:squaredlogerror, reg:logistic, reg:pseudohubererror, reg:tweedie, binary:logistic, binary:hinge, multi:softmax. Para obtener más información acerca de estos objetivos, consulte Learning task parameters en la documentación de XGBoost.

HYPERPARAMETERS { DEFAULT | DEFAULT EXCEPT ( key ‘value’ (,..) ) }

Especifica si los parámetros de XGBoost predeterminados se utilizan o se sustituyen con valores especificados por el usuario. Los valores deben ir entre comillas simples. Los siguientes son ejemplos de parámetros para XGBoost y sus valores predeterminados.

| Nombre del parámetro | Valor del parámetro | Valor predeterminado | Notas |
|---|---|---|---|
| num_class | Entero | Es obligatorio para la clasificación multiclase. | N/A |
| num_round | Entero | 100 | N/A |
| tree_method | Cadena | Automático | N/A |
| max_depth | Entero | 6 | [0, 10] |
| min_child_weight | Flotante | 1 | MinValue: 0, MaxValue: 120 |
| subsample | Flotante | 1 | MinValue: 0,5, MaxValue: 1 |
| gamma | Flotante | 0 | MinValue: 0, MaxValue: 5 |
| alpha | Flotante | 0 | MinValue: 0, MaxValue: 1000 |
| eta | Flotante | 0.3 | MinValue: 0,1, MaxValue: 0,5 |
| colsample_byleve | Flotante | 1 | MinValue: 0,1, MaxValue: 1 |
| colsample_bynode | Flotante | 1 | MinValue: 0,1, MaxValue: 1 |

| Nombre del parámetro | Valor del parámetro | Valor predeterminado | Notas |
|---|---|---|---|
| colsample_bytree | Flotante | 1 | MinValue: 0,5, MaxValue: 1 |
| lambda | Flotante | 1 | MinValue: 0, MaxValue: 1000 |
| max_delta_step | Entero | 0 | [0, 10] |

En el siguiente ejemplo, se preparan datos para XGBoost.

```
DROP TABLE IF EXISTS abalone_xgb;

CREATE TABLE abalone_xgb (
length_val float,
diameter float,
height float,
whole_weight float,
shucked_weight float,
viscera_weight float,
shell_weight float,
rings int,
record_number int);

COPY abalone_xgb
FROM 's3://redshift-downloads/redshift-ml/abalone_xg/'
REGION 'us-east-1'
IAM_ROLE default
IGNOREHEADER 1 CSV;
```

En el siguiente ejemplo, se crea un modelo XGBoost con las opciones avanzadas especificadas, como MODEL_TYPE, OBJECTIVE y PREPROCESSORS.

```
DROP MODEL abalone_xgboost_multi_predict_age;

CREATE MODEL abalone_xgboost_multi_predict_age
FROM ( SELECT length_val,
```

```
              diameter,
              height,
              whole_weight,
              shucked_weight,
              viscera_weight,
              shell_weight,
              rings
   FROM abalone_xgb WHERE record_number < 2500 )
TARGET rings FUNCTION ml_fn_abalone_xgboost_multi_predict_age
IAM_ROLE default
AUTO OFF
MODEL_TYPE XGBOOST
OBJECTIVE 'multi:softmax'
PREPROCESSORS 'none'
HYPERPARAMETERS DEFAULT EXCEPT (NUM_ROUND '100', NUM_CLASS '30')
SETTINGS (S3_BUCKET 'amzn-s3-demo-bucket');
```

En el siguiente ejemplo, se utiliza una consulta de inferencia para predecir la edad del pez con un número de registro superior a 2500. Utiliza la función ml_fn_abalone_xgboost_multi_predict_age creada a partir del comando anterior.

```
select ml_fn_abalone_xgboost_multi_predict_age(length_val,
                                                diameter,
                                                height,
                                                whole_weight,
                                                shucked_weight,
                                                viscera_weight,
                                                shell_weight)+1.5 as age
from abalone_xgb where record_number > 2500;
```

Bring your own model (BYOM): inferencia local

Amazon Redshift ML admite el uso de Bring your own model (BYOM) para inferencias locales.

A continuación, se resumen las opciones para la sintaxis de CREATE MODEL para BYOM. Puede utilizar un modelo entrenado fuera de Amazon Redshift con Amazon SageMaker AI para la inferencia local en la base de datos de Amazon Redshift.

Sintaxis de CREATE MODEL para inferencias locales

A continuación, se describe la sintaxis de CREATE MODEL para inferencias locales.

```
CREATE MODEL model_name
```

```
FROM ('job_name' | 's3_path' )
FUNCTION function_name ( data_type [, ...] )
RETURNS data_type
IAM_ROLE { default }
[ SETTINGS (
  S3_BUCKET 'amzn-s3-demo-bucket', | --required
  KMS_KEY_ID 'kms_string') --optional
];
```

Actualmente, Amazon Redshift solo admite modelos XGBoost, MLP y Linear Learner con formación previa para BYOM. Puede importar el Piloto automático de SageMaker AI y modelos entrenados directamente en Amazon SageMaker AI para inferencias locales con esta ruta.

Parámetros de CREATE MODEL para inferencias locales

model_name

El nombre del modelo. El nombre del modelo en un esquema debe ser único.

FROM ('job_name' | 's3_path' )

El parámetro job_name utiliza un nombre de trabajo de Amazon SageMaker AI como entrada. El nombre del trabajo puede ser el nombre de un trabajo de entrenamiento de Amazon SageMaker AI o el nombre de un trabajo de Piloto automático de Amazon SageMaker AI. El trabajo debe crearse en la misma cuenta de AWS propietaria del clúster de Amazon Redshift. Para encontrar el nombre del trabajo, inicialice Amazon SageMaker AI. En el menú desplegable de Training (Formación), elija Training jobs (Trabajos de formación).

El parámetro 's3_path' especifica la ubicación de S3 del archivo de artefactos del modelo .tar.gz que se utilizará cuando se cree el modelo.

FUNCTION function_name ( data_type [, ...] )

Se trata del nombre de la función que se creará y los tipos de datos de los argumentos de entrada. Puede proporcionar un nombre de esquema.

RETURNS data_type

El tipo de datos del valor que la función devuelve.

IAM_ROLE { default | 'arn:aws:iam::<account-id>:role/<role-name>'}

Utilice la palabra clave predeterminada para que Amazon Redshift utilice el rol de IAM configurado como predeterminado y asociado al clúster cuando se ejecuta el comando CREATE MODEL.

Utilice el nombre de recurso de Amazon (ARN), de un rol de IAM que el clúster utiliza para la autenticación y la autorización.

SETTINGS ( S3_BUCKET 'amzn-s3-demo-bucket', | KMS_KEY_ID 'kms_string')

La cláusula S3_BUCKET especifica la ubicación de Amazon S3 que se utiliza para almacenar resultados intermedios.

(Opcional) La cláusula KMS_KEY_ID especifica si Amazon Redshift utiliza cifrado del lado del servidor con una clave de AWS KMS para proteger los datos en reposo. Los datos en tránsito están protegidos con la capa de conexión segura (SSL).

Para obtener más información, consulte CREATE MODEL con guía para el usuario.

Ejemplo de CREATE MODEL para inferencias locales

En el siguiente ejemplo, se crea un modelo que se entrenó previamente en Amazon SageMaker, AI fuera de Amazon Redshift. Dado que Amazon Redshift ML admite el tipo de modelo para la inferencia local, el siguiente comando CREATE MODEL crea una función que se puede utilizar localmente en Amazon Redshift. Puede proporcionar el nombre de un trabajo de entrenamiento de SageMaker AI.

```
CREATE MODEL customer_churn
FROM 'training-job-customer-churn-v4'
FUNCTION customer_churn_predict (varchar, int, float, float)
RETURNS int
IAM_ROLE default
SETTINGS (S3_BUCKET 'amzn-s3-demo-bucket');
```

Una vez que se creó el modelo, puede utilizar la función customer_churn_predict con los tipos de argumentos especificados para generar predicciones.

Bring your own model (BYOM): inferencia remota

Amazon Redshift ML también admite el uso de Bring your own model (BYOM) para inferencias remotas.

A continuación, se resumen las opciones para la sintaxis de CREATE MODEL para BYOM.

Sintaxis de CREATE MODEL para inferencias remotas

A continuación, se describe la sintaxis de CREATE MODEL para inferencias remotas.

```
CREATE MODEL model_name
FUNCTION function_name ( data_type [, ...] )
RETURNS data_type
SAGEMAKER 'endpoint_name'[:'model_name']
IAM_ROLE { default | 'arn:aws:iam::<account-id>:role/<role-name>' }
[SETTINGS (MAX_BATCH_ROWS integer)];
```

Parámetros de CREATE MODEL para inferencias remotas

model_name

El nombre del modelo. El nombre del modelo en un esquema debe ser único.

FUNCTION fn_name ( [data_type] [, ...] )

Se trata del nombre de la función y los tipos de datos de los argumentos de entrada. Consulte Tipos de datos para ver todos los tipos de datos compatibles. `Geography`, `geometry` y `hllsketch` no son compatibles.

También puede proporcionar un nombre de función en un esquema mediante una notación de dos partes, como `myschema.myfunction`.

RETURNS data_type

El tipo de datos del valor que la función devuelve. Consulte Tipos de datos para ver todos los tipos de datos compatibles. `Geography`, `geometry` y `hllsketch` no son compatibles.

SAGEMAKER 'endpoint_name'[:'model_name']

El nombre del punto de conexión de Amazon SageMaker AI. Si el nombre del punto de conexión apunta a un punto de conexión de múltiples modelos, agregue el nombre del modelo que se utilizará. El punto de conexión debe estar alojado en la misma región de AWS y Cuenta de AWS que el clúster de Amazon Redshift. Para buscar el punto de conexión, inicialice Amazon SageMaker AI. En el menú desplegable de Inference (Inferencia), elija Endpoints (Puntos de conexión).

IAM_ROLE { default | 'arn:aws:iam::<account-id>:role/<role-name>'}

Utilice la palabra clave predeterminada para que Amazon Redshift utilice el rol de IAM configurado como predeterminado y asociado al clúster cuando se ejecuta el comando CREATE MODEL. También puede especificar el ARN de un rol de IAM para usar ese rol.

MAX_BATCH_ROWS integer

El número máximo de filas que Amazon Redshift envía en una sola solicitud por lotes para una única invocación de SageMaker AI. Solo se admite para BYOM con inferencia remota. El número real de filas de un lote también depende del tamaño de la entrada, pero es inferior o igual a este valor. El valor mínimo de este parámetro es 1. El valor máximo es `INT_MAX` o 2 147 483 647. Este parámetro solo es necesario cuando el tipo de datos de entrada y de devolución son `SUPER`. El valor predeterminado es `INT_MAX` o 2 147 483 647.

Cuando el modelo se implementa en un punto de conexión de SageMaker AI, este servicio crea la información del modelo en Amazon Redshift. Luego, realiza la inferencia a través de la función externa. Puede utilizar el comando SHOW MODEL para ver la información del modelo en el clúster de Amazon Redshift.

Notas de uso de CREATE MODEL para inferencias remotas

Antes de utilizar CREATE MODEL para inferencias remotas, tenga en cuenta lo siguiente:

- El punto de conexión debe estar alojado en la misma cuenta de AWS propietaria del clúster de Amazon Redshift.
- Asegúrese de que el punto de conexión de Amazon SageMaker AI tenga suficientes recursos para admitir las llamadas de inferencia de Amazon Redshift o de que el punto de conexión de Amazon SageMaker AI se pueda escalar de forma automática.
- Si no utiliza el tipo de datos `SUPER` como entrada, el modelo solo acepta entradas en formato de valores separados por comas (CSV), que corresponde a un tipo de contenido de `text/CSV` en SageMaker AI.
- Si no utiliza el tipo de datos `SUPER` como entrada, la salida de los modelos es un único valor del tipo especificado al crear la función. La salida está en el formato de valores separados por comas (CSV) en un tipo de contenido de `text/CSV` en SageMaker AI. Los tipos de datos `VARCHAR` no pueden estar entre comillas y no pueden contener saltos de línea. Además, cada salida debe estar en una nueva línea.
- Los modelos aceptan valores nulos como cadenas vacías.
- Si el tipo de datos de entrada es `SUPER`, solo se admite un argumento de entrada.
- Si el tipo de datos de entrada es `SUPER`, el tipo de datos de devolución también debe ser `SUPER`.
- MAX_BATCH_ROWS es necesario cuando los tipos de datos de entrada y de devolución son SUPER.

- Cuando el tipo de datos de entrada es SUPER, el tipo de contenido de la invocación del punto de conexión es `application/json` cuando MAX_BATCH_ROWS es 1 o `application/jsonlines` en todos los demás casos.
- Cuando el tipo de datos de devolución es SUPER, el tipo de aceptación de la invocación del punto de conexión es `application/json` cuando MAX_BATCH_ROWS es 1 o `application/jsonlines` en todos los demás casos.

### Ejemplo de CREATE MODEL para inferencias remotas

En el siguiente ejemplo, se crea un modelo que utiliza un punto de conexión de SageMaker AI para generar predicciones. Asegúrese de que el punto de conexión se esté ejecutando para generar predicciones y especifique su nombre en el comando CREATE MODEL.

```
CREATE MODEL remote_customer_churn
FUNCTION remote_fn_customer_churn_predict (varchar, int, float, float)
RETURNS int
SAGEMAKER 'customer-churn-endpoint'
IAM_ROLE default;
```

El siguiente ejemplo crea un BYOM con inferencia remota con un modelo de lenguaje grande (LLM). Los LLM alojados en Amazon Sagemaker AI Jumpstart aceptan y devuelven el tipo de contenido `application/json` y admiten un único JSON por invocación. Los tipos de datos de entrada y de devolución deben ser SUPER y MAX_BATCH_ROWS debe establecerse a 1.

```
CREATE MODEL sample_super_data_model
FUNCTION sample_super_data_model_predict(super)
RETURNS super
SAGEMAKER 'sample_super_data_model_endpoint'
IAM_ROLE default
SETTINGS (MAX_BATCH_ROWS 1);
```

### CREATE MODEL con K-MEANS

Amazon Redshift admite el algoritmo K-Means que agrupa datos que no están etiquetados. Este algoritmo resuelve problemas en clústeres en los que desea descubrir agrupaciones en los datos. Los datos sin clasificar se agrupan y dividen en función de sus similitudes y diferencias.

### CREATE MODEL con sintaxis K-MEANS

```
CREATE MODEL model_name
```

```
FROM { table_name | ( select_statement ) }
FUNCTION function_name
IAM_ROLE { default | 'arn:aws:iam::<account-id>:role/<role-name>' }
AUTO OFF
MODEL_TYPE KMEANS
PREPROCESSORS 'string'
HYPERPARAMETERS DEFAULT EXCEPT ( K 'val' [, ...] )
SETTINGS (
  S3_BUCKET 'amzn-s3-demo-bucket',
  KMS_KEY_ID 'kms_string', |
    -- optional
  S3_GARBAGE_COLLECT on / off, |
    -- optional
  MAX_CELLS integer, |
    -- optional
  MAX_RUNTIME integer
    -- optional);
```

CREATE MODEL con parámetros K-MEANS

AUTO OFF

Desactiva la detección automática de CREATE MODEL de procesadores previos, algoritmos e hiperparámetros.

MODEL_TYPE KMEANS

Especifica el uso de KMEANS para formar el modelo.

PREPROCESSORS 'string'

Especifica ciertas combinaciones de procesadores previos para determinados conjuntos de columnas. El formato es una lista de columnSets y las transformaciones adecuadas que se aplicarán a cada conjunto de columnas. Amazon Redshift admite 3 procesadores previos de K-Means, a saber, StandardScaler, MinMax y NumericPassthrough. Si no desea aplicar ningún procesamiento previo para K-Means, elija NumericPassthrough explícitamente como transformador. Para obtener más información sobre cómo usar los transformadores admitidos, consulte CREATE MODEL con parámetros de guía para el usuario.

El algoritmo K-Means utiliza la distancia euclidiana para calcular la similitud. El procesamiento previo de los datos garantiza que las características del modelo se mantengan a la misma escala y produzcan resultados confiables.

HYPERPARAMETERS DEFAULT EXCEPT ( K 'val' [, ...] )

Especifica si se utilizan los parámetros K-Means. Cuando se utiliza el algoritmo K-Means debe especificar el parámetro K. Para obtener más información, consulte Hiperparámetros de k-means en la Guía para el desarrollador de Amazon SageMaker AI

En el siguiente ejemplo, se preparan datos para K-Means.

```
CREATE MODEL customers_clusters
FROM customers
FUNCTION customers_cluster
IAM_ROLE default
AUTO OFF
MODEL_TYPE KMEANS
PREPROCESSORS '[
{
  "ColumnSet": [ "*" ],
  "Transformers": [ "NumericPassthrough" ]
}
]'
HYPERPARAMETERS DEFAULT EXCEPT ( K '5' )
SETTINGS (S3_BUCKET 'amzn-s3-demo-bucket');

select customer_id, customers_cluster(...) from customers;
customer_id | customers_cluster
--------------------
12345            1
12346            2
12347            4
12348
```

CREATE MODEL con Forecast

Los modelos de previsión de Redshift ML utilizan Amazon Forecast para crear previsiones de series temporales precisas. Esto le permite utilizar datos históricos de un periodo de tiempo para hacer predicciones sobre eventos futuros. Los casos de uso comunes de Amazon Forecast incluyen el uso de datos de productos minoristas para decidir el precio del inventario, los datos de cantidades de fabricación para predecir la cantidad de un artículo que se debe pedir y los datos de tráfico web para la previsión de la cantidad de tráfico que es posible que reciba un servidor web.

Los límites de cuota de Amazon Forecast se aplican a los modelos de previsión de Amazon Redshift. Por ejemplo, el número máximo de previsiones es 100, pero es ajustable. Al eliminar un modelo de previsión, no se eliminan automáticamente los recursos asociados en Amazon Forecast. Si elimina un clúster de Redshift, también se eliminarán todos los modelos asociados.

Tenga en cuenta que los modelos de previsión actualmente solo están disponibles en las siguientes regiones:

- Este de EE. UU. (Ohio) (us-east-2)
- Este de EE. UU. (Norte de Virginia) (us-east-1)
- Oeste de EE. UU. (Oregón) (us-west-2)
- Asia Pacífico (Bombay) (ap-south-1)
- Asia-Pacífico (Seúl) (ap-northeast-2)
- Asia-Pacífico (Singapur) (ap-southeast-1)
- Asia-Pacífico (Sídney) (ap-southeast-2)
- Asia-Pacífico (Tokio) (ap-northeast-1)
- Europa (Fráncfort) (eu-central-1)
- Europa (Irlanda) (eu-west-1)

### CREATE MODEL con sintaxis de Forecast

```
CREATE [ OR REPLACE ] MODEL forecast_model_name
FROM { table_name | ( select_query ) }
TARGET column_name
IAM_ROLE { default | 'arn:aws:iam::<account-id>:role/<role-name>'}
AUTO ON
MODEL_TYPE FORECAST
SETTINGS (
  S3_BUCKET 'amzn-s3-demo-bucket',
  HORIZON integer,
  FREQUENCY forecast_frequency
  [PERCENTILES '0.1', '0.5', '0.9']
  )
```

CREATE MODEL con parámetros de Forecast

forecast_model_name

El nombre del modelo. El nombre del modelo debe ser único.

FROM { table_name | ( select_query ) }

Se trata del nombre de la tabla o la consulta que especifica los datos de formación. Esto puede ser una tabla existente en el sistema o una consulta SELECT compatible con Amazon Redshift entre paréntesis. El resultado de la tabla o consulta debe tener al menos tres columnas: (1) una columna varchar que especifique el nombre de la serie temporal. Cada conjunto de datos puede tener varias series temporales; (2) una columna de fecha y hora; y (3) la columna de destino para predecir. Esta columna de destino debe ser int o flotante. Si proporciona un conjunto de datos que tiene más de tres columnas, Amazon Redshift asume que todas las columnas adicionales forman parte de una serie temporal relacionada. Tenga en cuenta que las series temporales relacionadas deben ser de tipo int o flotante. Para obtener más información sobre series temporales relacionadas, consulte Uso de conjuntos de datos de series temporales relacionadas.

TARGET column_name

Se trata del nombre de la columna que se convierte en el objetivo de predicción. La columna debe existir en la cláusula FROM.

IAM_ROLE { default | 'arn:aws:iam::<account-id>:role/<role-name>' }

Utilice la palabra clave predeterminada para que Amazon Redshift utilice el rol de IAM configurado como predeterminado y asociado al clúster cuando se ejecuta el comando CREAT MODEL. También puede especificar un ARN de un rol de IAM para usar ese rol.

AUTO ON

Activa la detección automática de CREATE MODEL del algoritmo y de la selección de hiperparámetros. Si especifica esta opción al crear un modelo de previsión, se debe utilizar un AutoPredictor de previsión, en el que Amazon Forecast aplica las combinaciones óptimas de algoritmos a cada serie temporal del conjunto de datos.

MODEL_TYPE FORECAST

Especifica el uso de FORECAST para formar el modelo.

S3_BUCKET 'amzn-s3-demo-bucket'

Se trata del nombre del bucket de Amazon Simple Storage Service que creó anteriormente y que se utiliza para compartir datos de formación y artefactos entre Amazon Redshift y Amazon Forecast. Amazon Redshift crea una subcarpeta en este bucket antes de descargar los datos de formación. Una vez que se completa la formación, Amazon Redshift elimina la subcarpeta creada y su contenido.

HORIZON integer

El número máximo de predicciones que puede devolver el modelo de previsión. Una vez que el modelo esté formado, no podrá cambiar este entero.

FREQUENCY forecast_frequency

Especifica el nivel de detalle que desea que tengan las previsiones. Las opciones disponibles son `Y | M | W | D | H | 30min | 15min | 10min | 5min | 1min`. Obligatorio si se entrena un modelo de previsión.

Cadena de PERCENTILES

Es una cadena delimitada por comas que especifica los tipos de previsión que se utilizan para entrenar un predictor. Los tipos de previsión pueden ser cuantiles de 0,01 a 0,99, por incrementos de 0,01 o más. También puede especificar la previsión media con la media. Puede especificar un máximo de cinco tipos de previsión.

En el siguiente ejemplo se muestra cómo crear un modelo de previsión sencillo.

```
CREATE MODEL forecast_example
FROM forecast_electricity_
TARGET target
IAM_ROLE 'arn:aws:iam::<account-id>:role/<role-name>'
AUTO ON
MODEL_TYPE FORECAST
SETTINGS (S3_BUCKET 'amzn-s3-demo-bucket',
          HORIZON 24,
          FREQUENCY 'H',
          PERCENTILES '0.25,0.50,0.75,mean',
          S3_GARBAGE_COLLECT OFF);
```

Tras crear el modelo de previsión, puede crear una tabla nueva con los datos de predicción.

```
CREATE TABLE forecast_model_results as SELECT Forecast(forecast_example)
```

A continuación, puede consultar la nueva tabla para obtener predicciones.

```
SELECT * FROM forecast_model_results
```

# CREATE PROCEDURE

Crea un procedimiento almacenado nuevo o sustituye un procedimiento existente para la actual base de datos.

Para obtener más información y ejemplos, consulta Creación de procedimientos almacenados en Amazon Redshift.

## Privilegios necesarios

Debe haber obtenido permiso de una de las siguientes maneras para ejecutar CREATE OR REPLACE PROCEDURE:

- Para CREATE PROCEDURE:
  - Superusuario
  - Usuarios con privilegios CREATE y USAGE en el esquema donde se crea el procedimiento almacenado
- Para REPLACE PROCEDURE:
  - Superusuario
  - Propietario del procedimiento

## Sintaxis

```
CREATE [ OR REPLACE ] PROCEDURE sp_procedure_name
  ( [ [ argname ] [ argmode ] argtype [, ...] ] )
[ NONATOMIC ]
AS $$
  procedure_body
$$ LANGUAGE plpgsql
[ { SECURITY INVOKER | SECURITY DEFINER } ]
[ SET configuration_parameter { TO value | = value } ]
```

## Parámetros

OR REPLACE

Una cláusula que especifica que si ya existe un procedimiento con el mismo nombre y los mismos tipos de datos de argumento de entrada o firma que este, debe reemplazarse el procedimiento existente. Solo puede reemplazar un procedimiento por un nuevo procedimiento que defina un conjunto idéntico de tipos de datos.

Si define un procedimiento con el mismo nombre que un procedimiento existente, pero con una firma diferente, creará un nuevo procedimiento. En otras palabras, se sobrecarga el nombre del procedimiento. Para obtener más información, consulte Sobrecarga de nombres de procedimientos.

sp_procedure_name

El nombre del procedimiento. Si especifica un nombre de esquema (como **`myschema.myprocedure`**), se crea el procedimiento en el esquema especificado. De lo contrario, el procedimiento se crea en el esquema actual. Para obtener más información acerca de los nombres válidos, consulte Nombres e identificadores.

Se recomienda que agregue el prefijo `sp_` a todos los nombres de procedimientos almacenados. Amazon Redshift reserva el prefijo `sp_` para los nombres de procedimientos almacenados. Con el uso del prefijo `sp_`, se asegura de que el nombre del procedimiento almacenado no entre en conflicto con el nombre de una función o un procedimiento almacenado integrado de Amazon Redshift existente o futuro. Para obtener más información, consulte Nomenclatura de los procedimientos almacenados.

Puede definir más de un procedimiento con el mismo nombre si los tipos de datos para los argumentos de entrada, o firmas, son diferentes. En otras palabras, en este caso se sobrecarga el nombre del procedimiento. Para obtener más información, consulte Sobrecarga de nombres de procedimientos

[argname] [ argmode] argtype

Una lista de nombres de argumento, modos de argumento y tipos de datos. Solo se necesita el tipo de datos. El nombre y el modo son opcionales y es posible cambiar su posición.

El modo del argumento puede ser IN, OUT O INOUT. El valor predeterminado es IN.

Puede utilizar argumentos OU e INOUT para devolver uno o más valores de una llamada de procedimientos. Cuando hay argumentos o INOUT, la llamada de procedimientos devuelve una

fila de resultados que contiene n columnas, done n es el número total de argumentos OUT o INOUT.

Los argumentos INOUT son argumentos de entrada y salida al mismo tiempo. Los argumentos de entrada incluyen argumentos IN e INOUT, y los argumentos de salida incluyen argumentos OUT e INOUT.

Los argumentos OUT no se especifican como parte de la instrucción CALL. Especifique argumentos INOUT en la declaración CALL de procedimientos almacenada. Los argumentos INOUT pueden ser útiles al pasar y devolver valores de una llamada anidada y también al devolver un `refcursor`. Para obtener más información sobre los tipos `refcursor`, consulte Cursores.

Los tipos de datos de argumentos pueden ser cualquier tipo de datos estándar de Amazon Redshift. Además, un tipo de datos de argumentos puede ser `refcursor`.

Puede especificar un máximo de 32 argumentos de entrada y 32 argumentos de salida.

AS $$ procedure_body $$

Se trata de una construcción que contiene el procedimiento que se ejecutará. Las palabras claves literales $$ y $$ son obligatorias.

Amazon Redshift exige que se incluya la instrucción en el procedimiento mediante un formato llamado entrecomillado con símbolo de dólar. Cualquier elemento que se encuentre dentro de los signos se transmite exactamente como es. No necesita incluir en una secuencia de escape los caracteres especiales porque el contenido de la cadena se escribe literalmente.

Con el entrecomillado de dólar, se utilizan un par de signos de dólar ($$) para representar el inicio y el final de la instrucción que se va a ejecutar, tal y como se muestra en el siguiente ejemplo.

```
$$ my statement $$
```

De manera opcional, entre los signos de dólar de cada par, puede especificar una cadena para ayudar a identificar la instrucción. La cadena que usa debe ser la misma en el inicio y en el final de los pares de signos. La cadena distingue entre mayúsculas y minúsculas, y sigue las mismas restricciones que un identificador sin comillas, excepto que no puede contener signos de dólar. En el siguiente ejemplo, se usa la prueba de cadena.

```
$test$ my statement $test$
```

Esta sintaxis también es útil para el entrecomillado de dólar anidado. Para obtener más información acerca de los signos de dólar, consulte"Dollar-quoted String Constants" bajo Lexical Structure en la documentación de PostgreSQL.

procedure_body

Un conjunto de instrucciones PL/pgSQL válidas. Las instrucciones PL/pgSQL aumentan los comandos SQL con construcciones de procedimiento, entre las que se incluyen expresiones condicionales y bucles, para controlar el flujo lógico. Se pueden usar la mayoría de los comandos SQL en el cuerpo del procedimiento, incluidos lenguajes de modificación de datos (DML) como COPY, UNLOAD e INSERT y lenguajes de definición de datos (DDL) como CREATE TABLE. Para obtener más información, consulte Referencia del lenguaje PL/pgSQL.

LANGUAGE plpgsql

Un valor del lenguaje. Especifique `plpgsql`. Debe tener permiso para usar el lenguaje para utilizar `plpgsql`. Para obtener más información, consulte GRANT.

NONATOMIC

Crea el procedimiento almacenado en un modo de transacción no atómico. El modo NONATOMIC confirma automáticamente las instrucciones que incluye el procedimiento. Además, cuando se produce un error en el procedimiento NONATOMIC, el error no se vuelve a generar si se gestiona mediante un bloque de excepciones. Para obtener más información, consulte Administración de transacciones y RAISE.

Cuando defina un procedimiento almacenado como `NONATOMIC`, tenga en cuenta lo siguiente:

- Cuando anide llamadas a procedimientos almacenados, todos deben crearse en el mismo modo de transacción.
- Las opciones `SECURITY DEFINER` y `SET configuration_parameter` no se admiten cuando se crea un procedimiento en modo NONATOMIC.
- Cualquier cursor que esté abierto (explícita o implícitamente) se cierra automáticamente cuando se procesa una confirmación implícita. Por lo tanto, debe abrir una transacción explícita antes de iniciar un bucle de cursor para asegurarse de que ninguna instrucción SQL en la iteración del bucle se confirme de forma implícita.

SECURITY INVOKER | SECURITY DEFINER

La opción `SECURITY DEFINER` no se admite cuando se especifica `NONATOMIC`.

El modo de seguridad para el procedimiento determina los privilegios de acceso del procedimiento en el tiempo de ejecución. El procedimiento tiene que tener permiso para acceder a los objetos subyacentes de la base de datos.

Para el modo SECURITY INVOKER, el procedimiento usa los privilegios del usuario que llama al procedimiento. El usuario tiene que tener permisos explícitos sobre los objetos subyacentes de la base de datos. El valor predeterminado es SECURITY INVOKER.

Para el modo SECURITY DEFINER, el procedimiento usa los privilegios del propietario del procedimiento. El propietario del procedimiento se define como el usuario propietario del procedimiento en tiempo de ejecución, no necesariamente el usuario que definió inicialmente el procedimiento. El usuario que llama al procedimiento necesita privilegio de ejecución en el procedimiento pero no necesita ningún privilegio en los objetos subyacentes.

SET configuration_parameter { TO value | = value }

Estas opciones no se admiten cuando se especifica `NONATOMIC`.

La cláusula SET causa que el parámetro `configuration_parameter` especificado se establezca en el valor especificado cuando se especifica el procedimiento. Esta cláusula restaura a continuación `configuration_parameter` a su valor anterior cuando el procedimiento existe.

## Notas de uso

Si se creó un procedimiento almacenado mediante la opción SECURITY DEFINER, al invocar la función CURRENT_USER desde el procedimiento almacenado, Amazon Redshift devuelve el nombre de usuario del propietario del procedimiento almacenado.

## Ejemplos

> **Note**
>
> Si al ejecutar estos ejemplos se produce un error similar a:
>
> ```
> ERROR: 42601: [Amazon](500310) unterminated dollar-quoted string at or near "$$
> ```
>
> Consulte Información general de procedimientos almacenados en Amazon Redshift.

El siguiente ejemplo crea un procedimiento con dos parámetros de entrada.

```
CREATE OR REPLACE PROCEDURE test_sp1(f1 int, f2 varchar(20))
AS $$
DECLARE
  min_val int;
BEGIN
  DROP TABLE IF EXISTS tmp_tbl;
  CREATE TEMP TABLE tmp_tbl(id int);
  INSERT INTO tmp_tbl values (f1),(10001),(10002);
  SELECT INTO min_val MIN(id) FROM tmp_tbl;
  RAISE INFO 'min_val = %, f2 = %', min_val, f2;
END;
$$ LANGUAGE plpgsql;
```

> **Note**
>
> Al escribir procedimientos almacenados, sugerimos una práctica recomendada para proteger los valores confidenciales:
> No codifique ninguna información confidencial en la lógica de procedimientos almacenada. Por ejemplo, no asigne una contraseña de usuario en una instrucción CREATE USER del cuerpo de un procedimiento almacenado. Esto supone un riesgo de seguridad, porque los valores codificados de forma rígida se pueden registrar como metadatos de esquema en las tablas del catálogo. En cambio, pase valores confidenciales, como las contraseñas, como argumentos al procedimiento almacenado, mediante parámetros.
> Para obtener más información acerca de los procedimientos almacenados, consultePROCEDIMIENTO DE CREACIÓNyCreación de procedimientos almacenados en Amazon Redshift. Para obtener más información acerca de las tablas de catálogos, consulte Tablas de catálogos de sistema.

El siguiente ejemplo crea un procedimiento con un parámetro IN, un parámetro OUT y un parámetro INOUT.

```
CREATE OR REPLACE PROCEDURE test_sp2(f1 IN int, f2 INOUT varchar(256), out_var OUT
 varchar(256))
AS $$
DECLARE
  loop_var int;
BEGIN
  IF f1 is null OR f2 is null THEN
    RAISE EXCEPTION 'input cannot be null';
```

```
  END IF;
  DROP TABLE if exists my_etl;
  CREATE TEMP TABLE my_etl(a int, b varchar);
    FOR loop_var IN 1..f1 LOOP
        insert into my_etl values (loop_var, f2);
        f2 := f2 || '+' || f2;
    END LOOP;
  SELECT INTO out_var count(*) from my_etl;
END;
$$ LANGUAGE plpgsql;
```

En el siguiente ejemplo se crea un procedimiento que usa el parámetro `SECURITY DEFINER`. Este procedimiento se ejecuta con los privilegios del usuario propietario del procedimiento.

```
CREATE OR REPLACE PROCEDURE sp_get_current_user_definer()
AS $$
DECLARE curr_user varchar(250);
BEGIN
  SELECT current_user INTO curr_user;
  RAISE INFO '%', curr_user;
END;
$$ LANGUAGE plpgsql
SECURITY DEFINER;
```

En el siguiente ejemplo se crea un procedimiento que usa el parámetro `SECURITY INVOKER`. Este procedimiento se ejecuta con los privilegios del usuario que ejecuta el procedimiento.

```
CREATE OR REPLACE PROCEDURE sp_get_current_user_invoker()
AS $$
DECLARE curr_user varchar(250);
BEGIN
  SELECT current_user INTO curr_user;
  RAISE INFO '%', curr_user;
END;
$$ LANGUAGE plpgsql
SECURITY INVOKER;
```

# CREATE RLS POLICY

Crea una nueva política de seguridad de la fila para proporcionar acceso pormenorizado a los objetos de la base de datos.

Los superusuarios y los usuarios o roles que tienen el rol sys:secadmin pueden crear una política.

## Sintaxis

```
CREATE RLS POLICY policy_name
[ WITH (column_name data_type [, ...]) [ [AS] relation_alias ] ]
USING ( using_predicate_exp )
```

## Parámetros

policy_name

El nombre de la política de .

WITH (column_name data_type [, ...])

Especifica column_name y data_type referenciados en las columnas de tablas a las que se adjunta la política.

Puede omitir la cláusula WITH solo cuando la política de RLS no haga referencia a ninguna columna de la tabla a la que se adjunta la política.

AS relation_alias

Especifica un alias opcional para la tabla a la que se adjuntará la política de RLS.

USING ( using_predicate_exp )

Especifica un filtro que se aplica a la cláusula WHERE de la consulta. Amazon Redshift aplica un predicado de política antes de los predicados de usuario de la consulta. Por ejemplo, **`current_user = 'joe' and price > 10`** limita a Joe a ver solo registros con un precio superior a 10 USD.

## Notas de uso

Al trabajar con la instrucción CREATE RLS POLICY, observe lo siguiente:

- Amazon Redshift admite filtros que pueden formar parte de la cláusula WHERE de una consulta.
- Todas las políticas que se adjuntan a una tabla deben haberse creado con el mismo alias de tabla.
- No necesita el permiso SELECT en las tablas de búsqueda. Al crear una política, Amazon Redshift concede el permiso SELECT en la tabla de búsqueda para la política correspondiente. Una tabla de búsqueda es un objeto de tabla que se utiliza dentro de la definición de una política.
- La seguridad de la fila de Amazon Redshift no admite los siguientes tipos de objeto dentro de una definición de política: tablas de catálogo, relaciones entre bases de datos, tablas externas, vistas normales, vistas de enlace en tiempo de ejecución, tablas con políticas de RLS activadas y tablas temporales.

## Ejemplos

Las siguientes instrucciones SQL crean las tablas, los usuarios y los roles para el ejemplo CREATE RLS POLICY.

```
-- Create users and roles reference in the policy statements.
CREATE ROLE analyst;

CREATE ROLE consumer;

CREATE USER bob WITH PASSWORD 'Name_is_bob_1';

CREATE USER alice WITH PASSWORD 'Name_is_alice_1';

CREATE USER joe WITH PASSWORD 'Name_is_joe_1';

GRANT ROLE sys:secadmin TO bob;

GRANT ROLE analyst TO alice;

GRANT ROLE consumer TO joe;

GRANT ALL ON TABLE tickit_category_redshift TO PUBLIC;
```

En el siguiente ejemplo, se crea una política denominada policy_concerts.

```
CREATE RLS POLICY policy_concerts
WITH (catgroup VARCHAR(10))
USING (catgroup = 'Concerts');
```

# CREAR ROL

Crea un nuevo rol personalizado que es un conjunto de permisos. Para obtener una lista de roles definidos por el sistema de Amazon Redshift, consulte the section called “Roles definidos por el sistema de Amazon Redshift”. Consulte SVV_ROLES para ver los roles creados actualmente en su clúster o grupo de trabajo.

Hay una cuota del número de roles que se pueden crear. Para obtener más información, consulte Cuotas y límites de Amazon Redshift en la Guía de administración de Amazon Redshift.

## Permisos necesarios

Los siguientes privilegios son necesarios para CREATE ROLE.

- Superusuario
- Usuarios con el privilegio CREATE ROLE

## Sintaxis

```
CREATE ROLE role_name
[ EXTERNALID external_id ]
```

## Parámetros

role_name

Nombre del rol El nombre del rol debe ser único y no puede ser el mismo que ningún nombre de usuario. El nombre de un rol no puede ser una palabra reservada.

Un superusuario o un usuario normal con el privilegio CREATE ROLE pueden crear roles. Un usuario que no sea superusuario pero al que se le haya concedido USAGE para el rol WITH GRANT OPTION y el privilegio ALTER puede conceder este rol a cualquiera.

EXTERNALID external_id

El identificador del rol, que está asociado a un proveedor de identidades. Para obtener más información, consulte Federación de proveedores de identidades (IdP) nativos para Amazon Redshift.

### Ejemplos

El siguiente ejemplo crea un rol `sample_role1`.

```
CREATE ROLE sample_role1;
```

En el ejemplo siguiente se crea un rol `sample_role1`, con un nuevo ID externo asociado a un proveedor de identidades.

```
CREATE ROLE sample_role1 EXTERNALID "ABC123";
```

## CREATE SCHEMA

Define un nuevo esquema para la base de datos actual.

### Privilegios necesarios

Los siguientes privilegios son necesarios para CREATE SCHEMA:

- Superusuario
- Usuarios con el privilegio CREATE SCHEMA

### Sintaxis

```
CREATE SCHEMA [ IF NOT EXISTS ] schema_name [ AUTHORIZATION username ]
           [ QUOTA {quota [MB | GB | TB] | UNLIMITED} ] [ schema_element [ ... ]

CREATE SCHEMA AUTHORIZATION username[ QUOTA {quota [MB | GB | TB] | UNLIMITED} ]
 [ schema_element [ ... ] ]
```

### Parámetros

IF NOT EXISTS

Cláusula que indica que si el esquema especificado ya existe, el comando no debe realizar cambios y debe devolver un mensaje en el que se indique que el esquema existe, en lugar de terminar con un error.

Esta cláusula es útil cuando se realiza scripting, para que el script no produzca un error si CREATE SCHEMA intenta crear un esquema que ya existe.

schema_name

Nombre del nuevo esquema. El nombre del esquema no puede ser `PUBLIC`. Para obtener más información acerca de los nombres válidos, consulte Nombres e identificadores.

> **Note**
>
> La lista de esquemas en el parámetro de configuración search_path determina la prioridad de objetos con nombres idénticos cuando se les hace referencia sin nombres de esquema.

AUTHORIZATION

Cláusula que otorga propiedad a un usuario especificado.

username

Nombre del propietario del esquema.

schema_element

Definición de uno o más objetos que se crearán dentro del esquema.

QUOTA

La cantidad máxima de espacio en disco que puede utilizar el esquema especificado. Este espacio es el uso colectivo del disco. Incluye todas las tablas permanentes, vistas materializadas bajo el esquema especificado y copias duplicadas de todas las tablas con distribución ALL en cada nodo de computación. La cuota de esquema no tiene en cuenta las tablas temporales creadas como parte de un espacio de nombres o un esquema temporales.

Para ver las cuotas del esquema configuradas, consulte SVV_SCHEMA_QUOTA_STATE.

Para ver los registros en los que se superaron las cuotas del esquema, consulte STL_SCHEMA_QUOTA_VIOLATIONS.

Amazon Redshift convierte el valor seleccionado en megabytes. Gigabytes es la unidad de medida predeterminada cuando no se especifica un valor.

Debe ser un superusuario de base de datos para establecer y cambiar una cuota de esquema. Un usuario que no sea superusuario pero que tenga permiso CREATE SCHEMA puede crear un esquema con una cuota definida. Cuando se crea un esquema sin definir una cuota, el esquema tiene una cuota ilimitada. Cuando se establece la cuota por debajo del valor actual utilizado por el esquema, Amazon Redshift no permite más ingesta hasta que se libere espacio en el disco. Una instrucción DELETE elimina datos de una tabla y el espacio en disco solo se libera cuando se ejecuta VACUUM.

Amazon Redshift verifica cada transacción en busca de infracciones de cuota antes de confirmar la transacción. Amazon Redshift revisa el tamaño (el espacio en disco utilizado por todas las tablas de un esquema) de cada esquema modificado en comparación con la cuota establecida. Dado que la comprobación de infracción de cuota se produce al final de una transacción, el límite de tamaño puede superar temporalmente la cuota dentro de una transacción antes de que se confirme. Cuando una transacción supera la cuota, Amazon Redshift detiene la transacción, prohíbe ingestas posteriores y revierte todos los cambios hasta que se libere espacio en el disco. Como la operación VACUUM y la limpieza interna se producen en segundo plano, es posible que un esquema no esté lleno al momento de su verificación después de que se cancela una transacción.

Como excepción, Amazon Redshift no tiene en cuenta la infracción de cuota y confirma transacciones en determinados casos. Amazon Redshift hace esto para transacciones que constan únicamente de una o más de las siguientes instrucciones donde no hay una instrucción de ingesta INSERT o COPY en la misma transacción:

- DELETE
- TRUNCATE
- VACUUM
- DROP TABLE
- ALTER TABLE APPEND solo cuando se mueven datos desde el esquema completo a otro esquema no completo

UNLIMITED

Amazon Redshift no impone ningún límite al crecimiento del tamaño total del esquema.

## Límites

Amazon Redshift aplica los siguientes límites para los esquemas.

- Hay un máximo de 9 900 esquemas por base de datos.

### Ejemplos

En el siguiente ejemplo, se crea un esquema denominado US_SALES y le otorga la propiedad al usuario DWUSER.

```
create schema us_sales authorization dwuser;
```

En el siguiente ejemplo, se crea un esquema denominado US_SALES, se otorga la propiedad al usuario DWUSER y se establece la cuota en 50 GB.

```
create schema us_sales authorization dwuser QUOTA 50 GB;
```

Para ver el nuevo esquema, consulte la tabla de catálogo PG_NAMESPACE, como se muestra a continuación.

```
select nspname as schema, usename as owner
from pg_namespace, pg_user
where pg_namespace.nspowner = pg_user.usesysid
and pg_user.usename ='dwuser';

   schema |  owner
----------+----------
 us_sales | dwuser
(1 row)
```

En el siguiente ejemplo, se crea el esquema US_SALES o no se hace nada y se devuelve un mensaje si el esquema ya existe:

```
create schema if not exists us_sales;
```

## CREATE TABLE

Crea una nueva tabla en la base de datos actual. Defina una lista de columnas, cada una de las cuales contiene datos de un tipo distinto. El propietario de la tabla es el emisor del comando CREATE TABLE.

## Privilegios necesarios

Los siguientes privilegios son necesarios para CREATE TABLE:

- Superusuario
- Usuarios con el privilegio CREATE TABLE

## Sintaxis

```
CREATE [ [LOCAL ] { TEMPORARY | TEMP } ] TABLE
[ IF NOT EXISTS ] table_name
( { column_name data_type [column_attributes] [ column_constraints ]
  | table_constraints
  | LIKE parent_table [ { INCLUDING | EXCLUDING } DEFAULTS ] }
  [, ... ]  )
[ BACKUP { YES | NO } ]
[table_attributes]

where column_attributes are:
  [ DEFAULT default_expr ]
  [ IDENTITY ( seed, step ) ]
  [ GENERATED BY DEFAULT AS IDENTITY ( seed, step ) ]
  [ ENCODE encoding ]
  [ DISTKEY ]
  [ SORTKEY ]
  [ COLLATE CASE_SENSITIVE | COLLATE CASE_INSENSITIVE  ]

and column_constraints are:
  [ { NOT NULL | NULL } ]
  [ { UNIQUE  |  PRIMARY KEY } ]
  [ REFERENCES reftable [ ( refcolumn ) ] ]

and table_constraints  are:
  [ UNIQUE ( column_name [, ... ] ) ]
  [ PRIMARY KEY ( column_name [, ... ] )  ]
  [ FOREIGN KEY (column_name [, ... ] ) REFERENCES reftable [ ( refcolumn ) ]


and table_attributes are:
  [ DISTSTYLE { AUTO | EVEN | KEY | ALL } ]
  [ DISTKEY ( column_name ) ]
  [ [COMPOUND | INTERLEAVED ] SORTKEY ( column_name [,...]) |  [ SORTKEY AUTO ] ]
```

```
  [ ENCODE AUTO ]
```

## Parámetros

LOCAL

Opcional. Aunque esta palabra clave se acepta en la instrucción, no tiene efecto en Amazon Redshift.

TEMPORARY | TEMP

Palabra clave que crea una tabla temporal que solo se puede ver dentro de la sesión actual. La tabla se elimina automáticamente al final de la sesión en la que se creó. La tabla temporal puede tener el mismo nombre que una tabla permanente. La tabla temporal se crea en un esquema separado específico de la sesión. (No se puede especificar un nombre para este esquema). Este esquema temporal se convierte en el primer esquema en la ruta de búsqueda, por lo que la tabla temporal tiene prioridad sobre la tabla permanente a menos que califique el nombre de la tabla con el nombre del esquema para obtener acceso a la tabla permanente. Para obtener más información acerca de esquemas y prioridades, consulte search_path.

> **Note**
>
> De forma predeterminada, los usuarios de base de datos tienen permiso para crear tablas temporales por su pertenencia automática al grupo PUBLIC. Para denegarle este privilegio a un usuario, anule el privilegio TEMP del grupo PUBLIC y, luego, conceda explícitamente el privilegio TEMP solo a usuarios o grupos de usuarios específicos.

IF NOT EXISTS

Cláusula que indica que si la tabla especificada ya existe, el comando no debe realizar cambios y debe devolver un mensaje en el que se indique que la tabla existe, en lugar de detenerse con un error. Tenga en cuenta que la tabla existente puede ser completamente diferente a la que podría haberse creado; solo se compara el nombre de la tabla.

Esta cláusula es útil cuando se realiza scripting, para que el script no produzca un error si CREATE TABLE intenta crear una tabla que ya existe.

table_name

Nombre de la tabla que se creará.

> **Important**
>
> Si especifica un nombre de tabla que comienza con "#", la tabla se crea como una tabla temporal. A continuación, se muestra un ejemplo:
>
> ```
> create table #newtable (id int);
> ```
>
> También hace referencia a la tabla con el carácter “#”. Por ejemplo:
>
> ```
> select * from #newtable;
> ```

La longitud máxima del nombre de la tabla es de 127 bytes; los nombres más largos se truncan en 127 bytes. Puede usar caracteres multibyte UTF-8 de hasta un máximo de cuatro bytes. Amazon Redshift aplica una cuota para el número de tablas por clúster según el tipo de nodo, incluidas las tablas temporales definidas por el usuario y las tablas temporales creadas por Amazon Redshift durante el procesamiento de consultas o el mantenimiento del sistema. De manera opcional, puede clasificar el nombre de la tabla con el nombre de la base de datos y del esquema. En el siguiente ejemplo, el nombre de base de datos es `tickit`, el nombre de esquema es `public` y el nombre de tabla es `test`.

```
create table tickit.public.test (c1 int);
```

Si no existe la base de datos o el esquema, no se crea la tabla y la instrucción devuelve un error. No puede crear tablas o vistas en las bases de datos del sistema `template0`, `template1`, `padb_harvest` o `sys:internal`.

Si se proporciona un nombre de esquema, la tabla nueva se crea en ese esquema (suponiendo que el creador tiene acceso al esquema). El nombre de la tabla debe ser un nombre único para ese esquema. Si no se especifica un esquema, se crea la tabla a través del esquema de la base de datos actual. Si está creando una tabla temporal, no puede especificar un nombre de esquema, ya que las tablas temporales existen en un esquema especial.

Pueden existir varias tablas temporales con el mismo nombre y al mismo tiempo en la misma base de datos si se crean en sesiones separadas porque las tablas se asignan a diferentes esquemas. Para obtener más información acerca de los nombres válidos, consulte Nombres e identificadores.

column_name

Nombre de una columna que se creará en la tabla nueva. La longitud máxima del nombre de la columna es de 127 bytes; los nombres más largos se truncan en 127 bytes. Puede usar caracteres multibyte UTF-8 de hasta un máximo de cuatro bytes. La cantidad máxima de columnas que se pueden definir en una única tabla es 1 600. Para obtener más información acerca de los nombres válidos, consulte Nombres e identificadores.

> **Note**
>
> Si está creando una "tabla ancha", asegúrese de que la lista de columnas no supere los límites de ancho de las filas para los resultados intermedios durante la carga y el procesamiento de consultas. Para obtener más información, consulte Notas de uso.

data_type

El tipo de datos de la columna que se crea. Para las columnas CHAR y VARCHAR, puede usar la palabra clave MAX en lugar de declarar una longitud máxima. MAX establece la longitud máxima en 4096 bytes para CHAR o 65 535 bytes para VARCHAR. El tamaño máximo del objeto GEOMETRY es 1 048 447 bytes.

Para obtener información acerca de los tipos de datos compatibles con Amazon Redshift, consulte Tipos de datos.

DEFAULT default_expr

Cláusula que asigna un valor de datos predeterminado para la columna. El tipo de datos de default_expr debe coincidir con el tipo de datos de la columna. El valor DEFAULT debe ser una expresión sin variables. No se permiten subconsultas, referencias cruzadas a otras columnas de la tabla actual ni funciones definidas por el usuario.

La expresión default_expr se utiliza en las operaciones INSERT que no especifican un valor para la columna. Si no se especifica un valor predeterminado, el valor predeterminado para la columna es nulo.

Si una operación COPY con una lista de columnas definida omite una columna que tiene el valor DEFAULT, el comando COPY inserta el valor de default_expr (expresión_predeterminada).

IDENTITY(seed, step)

Cláusula que especifica que la columna es una columna IDENTITY. Las columnas IDENTITY contienen valores únicos generados automáticamente. El tipo de datos de una columna IDENTITY debe ser INT o BIGINT.

Cuando agrega filas utilizando la instrucción `INSERT` o `INSERT INTO [tablename] VALUES( )`, estos valores comienzan por el valor especificado como seed (valor de inicialización) y van aumentando según el número especificado en step (paso).

Cuando la tabla se carga mediante la instrucción `INSERT INTO [tablename] SELECT * FROM` o `COPY` , los datos se cargan en paralelo y se distribuyen a los sectores del nodo. Para asegurarse de que los valores de identidad sean únicos, Amazon Redshift omite una serie de valores cuando crea los valores de identidad. Los valores de identidad son únicos, pero es posible que el orden no coincida con el de los archivos de origen.

GENERATED BY DEFAULT AS IDENTITY(seed, step)

Cláusula que especifica que la columna es una columna IDENTITY predeterminada y permite asignar automáticamente un valor único a la columna. El tipo de datos de una columna IDENTITY debe ser INT o BIGINT. Cuando agrega filas sin valores, estos valores comienzan con el valor especificado como seed (valor de inicialización) y van aumentando según el número especificado como step (paso). Para obtener más información acerca de cómo se generan los valores, consulte IDENTITY.

Además, durante las operaciones INSERT, UPDATE o COPY, puede proporcionar un valor sin EXPLICIT_IDS. Amazon Redshift utiliza ese valor para insertarlo en la columna de identidad en lugar de utilizar el valor generado por el sistema. El valor puede ser un duplicado, un valor inferior al valor de inicialización (seed) o un valor comprendido entre los valores de paso (step). Amazon Redshift no verifica la singularidad de los valores de la columna. El hecho de que se proporcione un valor no afecta al siguiente valor generado por el sistema.

> **Note**
>
> Si necesita valores eclusivos en la columna, no agregue un valor duplicado. En su lugar, agregue un valor único que sea inferior al valor de inicialización (seed) o que esté comprendido entre los valores de paso (step).

Tenga en cuenta lo siguiente sobre las columnas de identidad predeterminadas:

- Las columnas de identidad predeterminadas son NOT NULL. No se pueden insertar valores NULL.
- Para insertar un valor generado en una columna de identidad predeterminada, utilice la palabra clave DEFAULT.

```
INSERT INTO tablename (identity-column-name) VALUES (DEFAULT);
```

- El hecho de anular los valores de una columna de identidad predeterminada no afecta al siguiente valor generado.
- No puede agregar una columna de identidad predeterminada con la instrucción ALTER TABLE ADD COLUMN.
- Puede anexar una columna de identidad predeterminada con la instrucción ALTER TABLE APPEND.

ENCODE encoding

La codificación de compresión de una columna. ENCODE AUTO es la opción predeterminada para las tablas. Amazon Redshift administra de forma automática la codificación de compresión para todas las columnas de la tabla. Si especifica la codificación de compresión para alguna columna de la tabla, esta ya no tendrá la opción ENCODE AUTO configurada. Amazon Redshift ya no administra automáticamente la codificación de compresión para todas las columnas de la tabla. Puede especificar la opción ENCODE AUTO para la tabla con el fin de permitir que Amazon Redshift administre automáticamente la codificación de compresión para todas las columnas de la tabla.

Amazon Redshift asigna de forma automática una codificación de compresión inicial a las columnas para las que no se especifica la codificación de compresión de la siguiente manera:

- Por defecto, se asigna una compresión RAW a todas las columnas de tablas temporales.
- A las columnas que están definidas como claves de ordenación se les asigna una compresión RAW.
- A las columnas que están definidas como tipos de datos BOOLEAN, REAL, DOUBLE PRECISION, GEOMETRY o GEOGRAPHY se les asigna una compresión RAW.
- A las columnas que se definen como SMALLINT, INTEGER, BIGINT, DECIMAL, DATE, TIME, TIMETZ, TIMESTAMP o TIMESTAMPTZ se les asigna la compresión AZ64.
- A las columnas que se definen como CHAR, VARCHAR o VARBYTE se les asigna la compresión LZO.

> **Note**
>
> Si no desea que una columna se comprima, especifique explícitamente la codificación RAW.

Se admiten los siguientes compression encodings (p. 68):

- AZ64
- BYTEDICT
- DELTA
- DELTA32K
- LZO
- MOSTLY8
- MOSTLY16
- MOSTLY32
- RAW (sin comprimir)
- RUNLENGTH
- TEXT255
- TEXT32K
- ZSTD

DISTKEY

Palabra clave que especifica que la columna es la clave de distribución de la tabla. Solo una columna de una tabla puede ser la clave de distribución. Puede usar la palabra clave DISTKEY después del nombre de una columna o como parte de la definición de la tabla utilizando la sintaxis DISTKEY (column_name). Los dos métodos tienen el mismo efecto. Para obtener más información, consulte el parámetro DISTSTYLE más adelante en este tema.

El tipo de datos de una columna de clave de distribución puede ser algunas de las siguientes opciones: BOOLEAN, REAL, DOUBLE PRECISION, SMALLINT, INTEGER, BIGINT, DECIMAL, DATE, TIME, TIMETZ, TIMESTAMP, TIMESTAMPTZ, CHAR o VARCHAR.

SORTKEY

Palabra clave que especifica que la columna es la clave de ordenación de la tabla. Cuando carga datos en la tabla, los datos se ordenan por una o más columnas que se designan como claves

de ordenación. Puede usar la palabra clave SORTKEY después de un nombre de columna para especificar una clave de ordenación de una única columna, o puede especificar una o varias columnas como columnas de clave de ordenación de la tabla utilizando la sintaxis SORTKEY (column_name (nombre_de_columna) [, ...]). Solo se crean claves de ordenación compuestas con esta sintaxis.

Puede definir un máximo de 400 columnas SORTKEY por tabla.

El tipo de datos de una columna de clave de ordenación puede ser alguna de las siguientes opciones: BOOLEAN, REAL, DOUBLE PRECISION, SMALLINT, INTEGER, BIGINT, DECIMAL, DATE, TIME, TIMETZ, TIMESTAMP, TIMESTAMPTZ, CHAR o VARCHAR.

COLLATE CASE_SENSITIVE | COLLATE CASE_INSENSITIVE

Se trata de una cláusula que especifica si la búsqueda o la comparación de cadenas en la columna es CASE_SENSITIVE o CASE_INSENSITIVE. El valor predeterminado es el mismo que la configuración actual de distinción entre mayúsculas y minúsculas de la base de datos.

Para encontrar información de la intercalación de bases de datos, utilice el siguiente comando:

```
SELECT db_collation();

db_collation
----------------
 case_sensitive
(1 row)
```

NOT NULL | NULL

NOT NULL indica que la columna no puede contener valores nulos. NULL, el valor predeterminado, especifica que la columna acepta valores nulos. Las columnas IDENTITY están declaradas NOT NULL por defecto.

UNIQUE

Palabra clave que especifica que la columna puede contener solo valores únicos. El comportamiento de la restricción única de la tabla es el mismo que el de las restricciones de columna, con la capacidad adicional de abarcar varias columnas. Para definir una restricción de tabla única, consulte la sintaxis UNIQUE ( column_name [, ... ] ).

> **Important**
>
> Las restricciones únicas son informativas y el sistema no fuerza su aplicación.

PRIMARY KEY

Palabra clave que especifica que la columna es la clave principal de la tabla. Solo una columna se puede definir como la clave principal al utilizar una definición de columna. Para definir una restricción de tabla con una clave principal de varias columnas, use la sintaxis PRIMARY KEY ( column_name [, ... ] ).

Identificar una columna como clave principal proporciona metadatos acerca del diseño del esquema. Una clave principal implica que otras tablas pueden contar con este conjunto de columnas como un identificador único para las filas. Una clave principal puede especificarse para una tabla, ya sea como restricción de columna o restricción de tabla. La restricción de clave principal debe designar un conjunto de columnas diferente a los otros conjuntos de columnas denominados por cualquier restricción única definida para la misma tabla.

Las columnas CLAVE PRINCIPAL también se definen como NOT NULL.

> **Important**
>
> Las restricciones de clave principal son solo con fines informativos. El sistema no fuerza su aplicación, pero el planificador las utiliza.

References reftable [ ( refcolumn ) ]

Cláusula que especifica una restricción de clave externa, que implica que la columna debe contener solo los valores que coinciden con los valores en la columna referenciada de alguna fila de la tabla referenciada. Las columnas referenciadas deben ser las columnas de una clave única o principal en la tabla referenciada.

> **Important**
>
> Las restricciones de clave externa son solo con fines informativos. El sistema no fuerza su aplicación, pero el planificador las utiliza.

LIKE parent_table [ { INCLUDING | EXCLUDING } DEFAULTS ]

Una cláusula que especifica una tabla existente desde la cual la nueva tabla copia automáticamente nombres de columna, tipos de datos y restricciones NOT NULL. La tabla nueva y la tabla principal están desacopladas, por lo que los cambios realizados en la tabla principal no se aplicarán a la tabla nueva Las expresiones predeterminadas para las definiciones de columnas copiadas se copian solo si INCLUDING DEFAULTS está especificado. El comportamiento predeterminado es excluir las expresiones predeterminadas, para que todas las columnas de la tabla nueva tengan valores predeterminados nulos.

Las tablas creadas con la opción LIKE no heredan restricciones de clave principal y externa. Las tablas LIKE heredan el estilo de distribución, las claves de ordenación y las propiedades BACKUP y NULL, pero no se pueden configurar de forma explícita en la instrucción CREATE TABLE … LIKE.

BACKUP { YES | NO }

Una cláusula que especifica si la tabla debe incluirse en instantáneas de clústeres manuales y automáticos.

Para las tablas, como las tablas provisionales que no contienen datos críticos, especifique BACKUP NO para ahorrar tiempo de procesamiento al momento de crear instantáneas y restaurar contenido a partir de ellas, y para reducir espacio de almacenamiento en Amazon Simple Storage Service. El ajuste BACKUP NO no afecta la replicación automática de datos a otros nodos dentro del clúster, por lo que las tablas con BACKUP NO especificado se restauran en el caso de un error del nodo. El valor predeterminado es BACKUP YES.

> **Note**
>
> Las tablas sin copia de seguridad no son compatibles con RA3 y sin servidor. Una tabla marcada como sin copia de seguridad en RA3 y sin servidor se tratará como una tabla permanente de la que siempre se hará una copia de seguridad al tomar una instantánea, y que se restaurará al restaurar desde una instantánea.

DISTSTYLE { AUTO | EVEN | KEY | ALL }

Se trata de la palabra clave que define el estilo de distribución de datos para toda la tabla. Amazon Redshift distribuye las filas de una tabla entre los nodos informáticos en función del estilo de distribución especificado en la tabla. El valor predeterminado es AUTO.

El estilo de distribución que selecciona para las tablas afecta el rendimiento global de la base de datos. Para obtener más información, consulte Distribución de datos para la optimización de consultas. Los posibles estilos de distribución son los siguientes:

- AUTO: Amazon Redshift asigna un estilo de distribución óptimo en función de los datos de la tabla. Por ejemplo, si se especifica el estilo de distribución AUTO, Amazon Redshift asigna inicialmente el estilo de distribución ALL a una tabla pequeña. Cuando la tabla crezca, Amazon Redshift podría cambiar el estilo de distribución a KEY y elegir la clave principal (o una columna de la clave primaria compuesta) como DISTKEY. Si la tabla aumenta de tamaño y ninguna de las columnas es adecuada para ser DISTKEY, Amazon Redshift cambia el estilo de distribución a EVEN. El cambio en el estilo de distribución se produce en segundo plano y tiene un impacto mínimo en las consultas de los usuarios.

  Para ver el estilo de distribución aplicado a una tabla, consulte la tabla de catálogo del sistema PG_CLASS. Para obtener más información, consulte Visualización de los estilos de distribución.
- EVEN: Los datos de la tabla se distribuyen de manera uniforme en los nodos de un clúster en una distribución de turnos rotativos. Los ID de filas se utilizan para determinar la distribución, y se distribuye casi la misma cantidad de filas a cada nodo.
- KEY: Los datos se distribuyen por los valores en la columna DISTKEY. Cuando establece las columnas de unión de las tablas de unión como claves de distribución, las filas de combinación de ambas tablas se colocan en los nodos de computación. Cuando se colocan los datos, el optimizador puede realizar combinaciones de manera más eficiente. Si especifica DISTSTYLE KEY, debe nombrar una columna DISTKEY, ya sea para la tabla o como parte de la definición de la columna. Para obtener más información, consulte el parámetro DISTKEY mencionado en este tema.
- ALL: Una copia de toda la tabla se distribuye a cada nodo. El estilo de distribución garantiza que todas las filas obligatorias para cualquier combinación estén disponibles en todos los nodos, pero multiplica los requisitos de almacenamiento y aumenta el tiempo de carga y mantenimiento de la tabla. La distribución ALL puede mejorar el tiempo de ejecución cuando se usa con determinadas tablas de dimensión en las que la distribución KEY no es adecuada. No obstante, es conveniente analizar las mejoras de rendimiento en relación con los costos de mantenimiento.

DISTKEY ( column_name )

Restricción que especifica la columna que se utilizará como la clave de distribución de la tabla. Puede usar la palabra clave DISTKEY detrás del nombre de columna o como parte de la

definición de tabla utilizando la sintaxis DISTKEY (column_name). Los dos métodos tienen el mismo efecto. Para obtener más información, consulte el parámetro DISTSTYLE mencionado en este tema.

[COMPOUND | INTERLEAVED ] SORTKEY ( column_name [,...]) | [ SORTKEY AUTO ]

Especifica una o más claves de ordenación para la tabla. Cuando carga datos en la tabla, los datos se ordenan por las columnas que se designan como claves de ordenación. Puede usar la palabra clave SORTKEY después del nombre de una columna para especificar una clave de ordenación de una única columna, o puede especificar una o más columnas como las columnas de clave de ordenación para la tabla a través de la sintaxis `SORTKEY (column_name [ , ... ] )`.

Puede especificar el estilo de ordenación COMPOUND o INTERLEAVED. Si especifica SORTKEY con columnas, el valor predeterminado es COMPOUND. Para obtener más información, consulte Claves de clasificación.

Si no especifica ninguna opción de claves de ordenación, el valor predeterminado es AUTO.

Puede definir un máximo de 400 columnas COMPOUND SORTKEY u 8 columnas INTERLEAVED SORTKEY por tabla.

AUTO

Especifica que Amazon Redshift asigna una clave de ordenación óptima en función de los datos de la tabla. Por ejemplo, si se especifica la clave de ordenación AUTO, Amazon Redshift inicialmente no asigna ninguna clave de ordenación a una tabla. Si Amazon Redshift determina que una clave de ordenación mejorará el rendimiento de las consultas, dicho servicio podría cambiar la clave de ordenación de la tabla. La ordenación real de la tabla se realiza con la ordenación automática de la tabla. Para obtener más información, consulte Clasificación automática de tablas.

Amazon Redshift no modifica las tablas que ya tienen claves de ordenación o de distribución. Con una excepción, si una tabla tiene una clave de distribución que nunca se ha utilizado en una operación JOIN, la clave podría modificarse si Amazon Redshift determina que hay una clave mejor.

Para ver la clave de ordenación de una tabla, consulte la vista de catálogo del sistema SVV_TABLE_INFO. Para obtener más información, consulte SVV_TABLE_INFO. Si desea ver las recomendaciones para tablas de Advisor de Amazon Redshift, consulte

la vista de catálogo del sistema SVV_ALTER_TABLE_RECOMMENDATIONS. Para obtener más información, consulte SVV_ALTER_TABLE_RECOMMENDATIONS. Para ver las acciones llevadas a cabo por Amazon Redshift, consulte la vista de catálogo del sistema SVL_AUTO_WORKER_ACTION. Para obtener más información, consulte SVL_AUTO_WORKER_ACTION.

COMPOUND

Especifica que los datos se ordenan a través de una clave compuesta formada por todas las columnas enumeradas en el orden en que aparecen en la lista. Una clave de ordenación compuesta es útil cuando una consulta analiza filas según el orden de las columnas de ordenación. Los beneficios del rendimiento de ordenación con una clave compuesta se reducen cuando las consultas dependen de columnas de ordenación secundarias. Puede definir una cantidad máxima de 400 columnas COMPOUND SORTKEY por tabla.

INTERLEAVED

Especifica que los datos se ordenan a través de una clave de ordenación intercalada. Se puede especificar una cantidad máxima de ocho columnas para una clave de ordenación intercalada.

Una ordenación intercalada otorga el mismo peso a cada columna o subconjunto de columnas de la clave de ordenación. Por lo tanto, las consultas no dependen del orden que tengan las columnas en la clave de ordenación. Cuando una consulta usa una o más columnas de ordenación secundarias, la ordenación intercalada mejora considerablemente el rendimiento de la consulta. La ordenación intercalada tiene un pequeño costo general para las operaciones de carga y limpieza de datos.

> **Important**
>
> No utilice una clave de ordenación intercalada con atributos monótonamente crecientes, como columnas de identidad, fechas o marcas temporales.

ENCODE AUTO

Permite a Amazon Redshift ajustar de forma automática el tipo de codificación de todas las columnas de la tabla para optimizar el rendimiento de la consulta. ENCODE AUTO conserva los tipos de codificación iniciales que se especifican durante la creación de la tabla. Luego, si Amazon Redshift determina que un nuevo tipo de codificación puede mejorar el rendimiento

de la consulta, Amazon Redshift puede cambiar el tipo de codificación de las columnas de la tabla. ENCODE AUTO es la opción predeterminada si no se especifica un tipo de codificación en cualquiera de las columnas de la tabla.

UNIQUE ( column_name [,...] )

Restricción que especifica que un grupo de una o más columnas de una tabla puede contener solo valores únicos. El comportamiento de la restricción única de la tabla es el mismo que el de las restricciones de columna, con la capacidad adicional de abarcar varias columnas. En el contexto de las restricciones únicas, los valores nulos no se consideran iguales. Cada restricción de tabla única debe designar un conjunto de columnas diferente al conjunto de columnas denominado por otra restricción de clave única o principal definida para la tabla.

> **Important**
>
> Las restricciones únicas son informativas y el sistema no fuerza su aplicación.

PRIMARY KEY ( column_name [,...] )

Restricción que especifica que una columna o un grupo de columnas de una tabla puede contener solo valores no nulos (no duplicados). Identificar un conjunto de columnas como clave principal también proporciona metadatos acerca del diseño del esquema. Una clave principal implica que otras tablas pueden contar con este conjunto de columnas como un identificador único para las filas. Una clave principal puede especificarse para una tabla, ya sea como restricción de columna única o restricción de tabla. La restricción de clave principal debe designar un conjunto de columnas diferente a los otros conjuntos de columnas denominados por cualquier restricción única definida para la misma tabla.

> **Important**
>
> Las restricciones de clave principal son solo con fines informativos. El sistema no fuerza su aplicación, pero el planificador las utiliza.

FOREIGN KEY ( column_name [, ... ] ) REFERENCES reftable [ ( refcolumn ) ]

Restricción que especifica una restricción de clave externa, que requiere que un grupo de una o más columnas de la nueva tabla debe contener solo valores que coincidan con los valores en la

columna o columnas a las que se hace referencia de alguna fila de la tabla referenciada. Si se omite refcolumn , se usa la clave principal de reftable. Las columnas referenciadas deben ser las columnas de una clave única o principal en la tabla referenciada.

> **Important**
>
> Las restricciones de clave externa son solo con fines informativos. El sistema no fuerza su aplicación, pero el planificador las utiliza.

## Notas de uso

Las restricciones de unicidad, clave principal y clave externa son solo informativas, Amazon Redshift no las aplica al rellena una tabla. Por ejemplo, si inserta datos en una tabla con dependencias, la inserción puede realizarse correctamente aunque infrinja la restricción. No obstante, las claves principales y externas se utilizan como sugerencias de planificación y deben estar declaradas si el proceso de Extract, Transform, Load (ETL, Extracción, transformación y carga) o algún otro proceso de su aplicación exige su integridad. Para obtener información acerca de cómo eliminar una tabla con dependencias, consulte DROP TABLE.

### Límites y cuotas

Tenga en cuenta los siguientes límites al crear una tabla.

- Existe un límite para el número máximo de tablas en un clúster por tipo de nodo. Para obtener más información, consulte Límites en la Guía de administración de Amazon Redshift.
- La cantidad máxima de caracteres para el nombre de una tabla es 127.
- La cantidad máxima de columnas que se pueden definir en una única tabla es 1 600.
- La cantidad máxima de columnas SORTKEY que se pueden definir en una única tabla es 400.

### Resumen de ajustes de nivel de columna y ajustes de nivel de tabla

Se pueden configurar varios atributos y ajustes al nivel de columna o al nivel de tabla. En algunos casos, configurar un atributo o una restricción al nivel de columna o al nivel de tabla tiene el mismo efecto. En otros casos, producen resultados diferentes.

La siguiente lista resume los ajustes de nivel de columna y de nivel de tabla:

DISTKEY

No hay diferencia en el efecto si se configura al nivel de columna o al nivel de tabla.

Si se establece DISTKEY, ya sea al nivel de columna o al nivel de tabla, DISTSTYLE debe estar configurado en KEY o no estar configurado. DISTSTYLE solo puede configurarse al nivel de tabla.

SORTKEY

Si se configura al nivel de columna, SORTKEY debe ser una única columna. Si SORTKEY se configura al nivel de tabla, una o más columnas pueden formar una clave de ordenación compuesta o intercalada.

COLLATE CASE_SENSITIVE | COLLATE CASE_INSENSITIVE

Amazon Redshift no admite la modificación de la distinción entre mayúsculas y minúsculas de una columna. Cuando agrega una columna nueva a la tabla, Amazon Redshift utiliza el valor predeterminado para la distinción entre mayúsculas y minúsculas. Amazon Redshift no admite la palabra clave COLLATE cuando se anexa una columna nueva.

Para obtener información sobre cómo crear bases de datos mediante la intercalación de bases de datos, consulte CREATE DATABASE.

Para obtener información acerca de la función COLLATE, consulte Función COLLATE.

UNIQUE

Al nivel de columna, una o más claves pueden configurarse en UNIQUE. La restricción UNIQUE se aplica a cada columna individualmente. Si UNIQUE se configura al nivel de tabla, una o más columnas pueden formar una restricción UNIQUE compuesta.

PRIMARY KEY

Si se configura al nivel de columna, PRIMARY KEY debe ser una única columna. Si PRIMARY KEY se configura al nivel de tabla, una o más columnas pueden formar una clave principal compuesta.

FOREIGN KEY

No hay diferencia en el efecto si se configura FOREIGN KEY al nivel de columna o al nivel de tabla. Al nivel de columna, la sintaxis es `REFERENCES reftable [ ( refcolumn )]`.

Distribución de datos entrantes

Cuando el esquema de distribución hash de los datos entrantes coincide con el de la tabla de destino, no es necesaria la distribución física de datos cuando se cargan los datos. Por ejemplo, si se configura una clave de distribución para la nueva tabla y los datos se insertan desde otra tabla que está distribuida en la misma columna clave, los datos se cargan en el lugar a través de los mismos nodos y sectores. No obstante, si las tablas de origen y destino se configuran en distribución EVEN, los datos se redistribuyen en la tabla de destino.

Tablas anchas

Puede crear una tabla muy ancha pero no tener la capacidad de realizar el procesamiento de consultas, como instrucciones INSERT o SELECT, en la tabla. El ancho máximo de una tabla con columnas de ancho fijo, como CHAR, es 64 KB - 1 (o 65 535 bytes). Si una tabla contiene columnas VARCHAR, la tabla puede tener un ancho declarado superior sin devolver un error, ya que, en las columnas VARCHARS, el ancho declarado completo no afecta al límite calculado de procesamiento de consultas. El límite de procesamiento de consultas en vigor para las columnas VARCHAR variará en función a un número de factores.

Si la tabla es demasiado ancha para insertar o seleccionar, recibe el siguiente error.

```
ERROR:  8001
DETAIL:  The combined length of columns processed in the SQL statement
exceeded the query-processing limit of 65535 characters (pid:7627)
```

## Ejemplos

Para ver ejemplos que muestran cómo usar el comando CREATE TABLE, consulte el tema Ejemplos.

## Ejemplos

Los siguientes ejemplos muestran distintos atributos de tabla y columna en instrucciones CREATE TABLE de Amazon Redshift. Para obtener más información acerca de CREATE TABLE, incluidas las definiciones de parámetros, consulte CREATE TABLE.

Muchos de los ejemplos utilizan tablas y datos del conjunto de datos de muestra TICKIT. Para obtener más información, consulte Base de datos de muestra.

En un comando CREATE TABLE puede anteponer el nombre de base de datos y el nombre de esquema al nombre de tabla. Por ejemplo, `dev_database.public.sales`. El nombre de base de

datos debe ser la base de datos a la que está conectado. Cualquier intento de crear objetos de base de datos en otra base de datos produce un error de operación no válida.

Crear una tabla con una clave de distribución, una clave de ordenación compuesta y compresión

En el siguiente ejemplo, se crea una tabla SALES en la base de datos TICKIT con compresión definida para varias columnas. LISTID está declarada como la clave de distribución, y LISTID y SELLERID están declaradas como una clave de ordenación compuesta de varias columnas. También se definen las restricciones de clave principal y clave externa para la tabla. Antes de crear la tabla del ejemplo, puede que necesite agregar una restricción UNIQUE a cada columna a la que haga referencia una clave externa, si no existen restricciones.

```
create table sales(
salesid integer not null,
listid integer not null,
sellerid integer not null,
buyerid integer not null,
eventid integer not null encode mostly16,
dateid smallint not null,
qtysold smallint not null encode mostly8,
pricepaid decimal(8,2) encode delta32k,
commission decimal(8,2) encode delta32k,
saletime timestamp,
primary key(salesid),
foreign key(listid) references listing(listid),
foreign key(sellerid) references users(userid),
foreign key(buyerid) references users(userid),
foreign key(dateid) references date(dateid))
distkey(listid)
compound sortkey(listid,sellerid);
```

Los resultados son los siguientes:

```
schemaname | tablename | column     | type                          | encoding | distkey
 | sortkey | notnull
-----------+-----------+------------+-------------------------------+----------+---------
+---------+--------
public     | sales     | salesid    | integer                       | lzo      | false
 |       0 | true
public     | sales     | listid     | integer                       | none     | true
 |       1 | true
```

```
public     | sales     | sellerid   | integer                     | none     | false
 |       2 | true
public     | sales     | buyerid    | integer                     | lzo      | false
 |       0 | true
public     | sales     | eventid    | integer                     | mostly16 | false
 |       0 | true
public     | sales     | dateid     | smallint                    | lzo      | false
 |       0 | true
public     | sales     | qtysold    | smallint                    | mostly8  | false
 |       0 | true
public     | sales     | pricepaid  | numeric(8,2)                | delta32k | false
 |       0 | false
public     | sales     | commission | numeric(8,2)                | delta32k | false
 |       0 | false
public     | sales     | saletime   | timestamp without time zone | lzo      | false
 |       0 | false
```

En el siguiente ejemplo, se crea la tabla t1 con una columna col1 que no distingue entre mayúsculas y minúsculas.

```
create table T1 (
  col1 Varchar(20) collate case_insensitive
 );

insert into T1 values ('bob'), ('john'), ('Tom'), ('JOHN'), ('Bob');
```

Consultar la tabla:

```
select * from T1 where col1 = 'John';

col1
------
 john
 JOHN
(2 rows)
```

Crear una tabla con una clave de ordenación intercalada

En el siguiente ejemplo, se crea una tabla CUSTOMER con una clave de ordenación intercalada.

```
create table customer_interleaved (
  c_custkey     integer        not null,
```

```
  c_name         varchar(25)    not null,
  c_address      varchar(25)    not null,
  c_city         varchar(10)    not null,
  c_nation       varchar(15)    not null,
  c_region       varchar(12)    not null,
  c_phone        varchar(15)    not null,
  c_mktsegment      varchar(10)    not null)
diststyle all
interleaved sortkey (c_custkey, c_city, c_mktsegment);
```

Crear una tabla con IF NOT EXISTS

En el siguiente ejemplo, se crea la tabla CITIES o no se realiza una acción y se devuelve un mensaje si ya existe:

```
create table if not exists cities(
cityid integer not null,
city varchar(100) not null,
state char(2) not null);
```

Crear una tabla con la distribución ALL

En el siguiente ejemplo, se crea la tabla VENUE con la distribución ALL.

```
create table venue(
venueid smallint not null,
venuename varchar(100),
venuecity varchar(30),
venuestate char(2),
venueseats integer,
primary key(venueid))
diststyle all;
```

Crear una tabla con la distribución EVEN

En el siguiente ejemplo, se crea una tabla denominada MYEVENT con tres columnas.

```
create table myevent(
eventid int,
eventname varchar(200),
eventcity varchar(30))
diststyle even;
```

La tabla se distribuye de manera uniforme y no se ordena. La tabla no tiene columnas DISTKEY o SORTKEY declaradas.

```
select "column", type, encoding, distkey, sortkey
from pg_table_def where tablename = 'myevent';

  column   |          type          | encoding | distkey | sortkey
-----------+------------------------+----------+---------+---------
 eventid   | integer                | lzo      | f       |       0
 eventname | character varying(200) | lzo      | f       |       0
 eventcity | character varying(30)  | lzo      | f       |       0
(3 rows)
```

Crear una tabla temporal con el comando LIKE que sea como otra tabla

En el siguiente ejemplo, se crea una tabla temporal denominada TEMPEVENT, que hereda sus columnas de la tabla EVENT.

```
create temp table tempevent(like event);
```

Esta tabla también hereda los atributos DISTKEY y SORTKEY de su tabla principal:

```
select "column", type, encoding, distkey, sortkey
 from pg_table_def where tablename = 'tempevent';

  column   |            type             | encoding | distkey | sortkey
-----------+-----------------------------+----------+---------+---------
 eventid   | integer                     | none     | t       |       1
 venueid   | smallint                    | none     | f       |       0
 catid     | smallint                    | none     | f       |       0
 dateid    | smallint                    | none     | f       |       0
 eventname | character varying(200)      | lzo      | f       |       0
 starttime | timestamp without time zone | bytedict | f       |       0
(6 rows)
```

Crear una tabla con una columna IDENTITY

En el siguiente ejemplo, se crea una tabla denominada VENUE_IDENT, que tiene una columna IDENTITY denominada VENUEID. Esta columna comienza con 0 y crece en incrementos de 1 para cada registro. VENUEID también se declara como la clave principal de la tabla.

```
create table venue_ident(venueid bigint identity(0, 1),
```

```
venuename varchar(100),
venuecity varchar(30),
venuestate char(2),
venueseats integer,
primary key(venueid));
```

Crear una tabla con una columna IDENTITY predeterminada

En el siguiente ejemplo, se crea una tabla llamada `t1`. Esta tabla tiene una columna IDENTITY llamada `hist_id` y una columna IDENTITY predeterminada llamada `base_id`.

```
CREATE TABLE t1(
  hist_id BIGINT IDENTITY NOT NULL, /* Cannot be overridden */
  base_id BIGINT GENERATED BY DEFAULT AS IDENTITY NOT NULL, /* Can be overridden */
  business_key varchar(10) ,
  some_field varchar(10)
);
```

Si se inserta una fila en la tabla, se generan los dos valores: `hist_id` y `base_id`.

```
INSERT INTO T1 (business_key, some_field) values ('A','MM');
```

```
SELECT * FROM t1;

 hist_id | base_id | business_key | some_field
---------+---------+--------------+------------
       1 |       1 | A            | MM
```

Si se inserta una segunda fila, se genera el valor predeterminado de `base_id`.

```
INSERT INTO T1 (base_id, business_key, some_field) values (DEFAULT, 'B','MNOP');
```

```
SELECT * FROM t1;

 hist_id | base_id | business_key | some_field
---------+---------+--------------+------------
       1 |       1 | A            | MM
       2 |       2 | B            | MNOP
```

Si se inserta una tercera fila, no es necesario que el valor de `base_id` sea único.

```
INSERT INTO T1 (base_id, business_key, some_field) values (2,'B','MNNN');
```

```
SELECT * FROM t1;

 hist_id | base_id | business_key | some_field
---------+---------+--------------+------------
       1 |       1 | A            | MM
       2 |       2 | B            | MNOP
       3 |       2 | B            | MNNN
```

Crear una tabla con valores de la columna DEFAULT

En el siguiente ejemplo, se crea una tabla CATEGORYDEF que declara valores predeterminados para cada columna:

```
create table categorydef(
catid smallint not null default 0,
catgroup varchar(10) default 'Special',
catname varchar(10) default 'Other',
catdesc varchar(50) default 'Special events',
primary key(catid));

insert into categorydef values(default,default,default,default);
```

```
select * from categorydef;

 catid | catgroup | catname |    catdesc
-------+----------+---------+----------------
     0 | Special  | Other   | Special events
(1 row)
```

Opciones DISTSTYLE, DISTKEY y SORTKEY

En el siguiente ejemplo, se muestra cómo funcionan las opciones DISTKEY, SORTKEY y DISTSTYLE. En este ejemplo, COL1 es la clave de distribución: por lo tanto, el estilo de distribución debe configurarse en KEY o no configurarse. De manera predeterminada, la tabla no tiene ninguna clave de ordenación y, por lo tanto, no se ordena:

```
create table t1(col1 int distkey, col2 int) diststyle key;
```

```
select "column", type, encoding, distkey, sortkey
from pg_table_def where tablename = 't1';

column |  type   | encoding | distkey | sortkey
-------+---------+----------+---------+---------
col1   | integer | az64     | t       | 0
col2   | integer | az64     | f       | 0
```

En el siguiente ejemplo, se define la misma columna como la clave de distribución y la clave de ordenación. El estilo de distribución debe configurarse en KEY o no configurarse.

```
create table t2(col1 int distkey sortkey, col2 int);
```

```
select "column", type, encoding, distkey, sortkey
from pg_table_def where tablename = 't2';

column |  type   | encoding | distkey | sortkey
-------+---------+----------+---------+---------
col1   | integer | none     | t       | 1
col2   | integer | az64     | f       | 0
```

En el siguiente ejemplo, no se configura una columna como la clave de distribución, se configura COL2 como la clave de ordenación, y se configura el estilo de distribución en ALL:

```
create table t3(col1 int, col2 int sortkey) diststyle all;
```

```
select "column", type, encoding, distkey, sortkey
from pg_table_def where tablename = 't3';

Column |  Type   | Encoding | DistKey | SortKey
-------+---------+----------+---------+--------
col1   | integer | az64     | f       | 0
col2   | integer | none     | f       | 1
```

En el siguiente ejemplo, el estilo de distribución está establecido en EVEN y no se define explícitamente una clave de ordenación. Por lo tanto, la tabla se distribuye de manera uniforme, pero no se ordena.

```
create table t4(col1 int, col2 int) diststyle even;
```

```
select "column", type, encoding, distkey, sortkey
from pg_table_def where tablename = 't4';

             column |  type   |encoding | distkey | sortkey
--------+---------+---------+---------+--------
col1    | integer | az64    | f       | 0
col2    | integer | az64    | f       | 0
```

Crear una tabla con la opción ENCODE AUTO

En el siguiente ejemplo, se crea la tabla `t1` con codificación de compresión automática. ENCODE AUTO es la opción predeterminada para las tablas cuando no se especifica un tipo de codificación para ninguna columna.

```
create table t1(c0 int, c1 varchar);
```

En el siguiente ejemplo, se crea la tabla `t2` con codificación de compresión automática mediante la especificación de ENCODE AUTO.

```
create table t2(c0 int, c1 varchar) encode auto;
```

En el siguiente ejemplo, se crea la tabla `t3` con codificación de compresión automática mediante la especificación de ENCODE AUTO. La columna `c0` se define con un tipo de codificación inicial de DELTA. Amazon Redshift puede cambiar la codificación si otra codificación proporciona un mejor rendimiento en las consultas.

```
create table t3(c0 int encode delta, c1 varchar) encode auto;
```

En el siguiente ejemplo, se crea la tabla `t4` con codificación de compresión automática mediante la especificación de ENCODE AUTO. La columna `c0` se define con una codificación inicial de DELTA, y la columna `c1` se define con una codificación inicial de LZO. Amazon Redshift puede cambiar estas codificaciones si otras codificaciones proporcionan un mejor rendimiento en las consultas.

```
create table t4(c0 int encode delta, c1 varchar encode lzo) encode auto;
```

# CREATE TABLE AS

Temas

- Sintaxis
- Parámetros
- Notas de uso de CTAS
- Ejemplos de CTAS


Crea una nueva tabla en función de una consulta. El propietario de esta tabla es el usuario que emite el comando.

La tabla nueva se carga con datos definidos por la consulta en el comando. Las columnas de la tabla tienen nombres y tipos de datos asociados con las columnas de salida de la consulta. El comando CREATE TABLE AS (CTAS) crea una tabla nueva y evalúa la consulta para cargar la tabla nueva.

## Sintaxis

```
CREATE [ [ LOCAL ] { TEMPORARY | TEMP } ]
TABLE table_name
[ ( column_name [, ... ] ) ]
[ BACKUP { YES | NO } ]
[ table_attributes ]
AS query

where table_attributes are:
[ DISTSTYLE { AUTO | EVEN | ALL | KEY } ]
[ DISTKEY( distkey_identifier ) ]
[ [ COMPOUND | INTERLEAVED ] SORTKEY( column_name [, ...] ) ]
```

## Parámetros

LOCAL

Aunque esta palabra clave opcional se acepta en la instrucción, no tiene efecto en Amazon Redshift.

TEMPORARY | TEMP

Crea una tabla temporal. Una tabla temporal se elimina automáticamente al final de la sesión en la que se creó.

table_name

El nombre de la tabla que se creará.

> **Important**
>
> Si especifica un nombre de tabla que comienza con "#", la tabla se crea como una tabla temporal. Por ejemplo:
>
> ```
> create table #newtable (id) as select * from oldtable;
> ```

La longitud máxima del nombre de la tabla es de 127 bytes; los nombres más largos se truncan en 127 bytes. Amazon Redshift aplica una cuota del número de tablas por clúster según el tipo de nodo. Puede clasificar el nombre de la tabla con el nombre de la base de datos y del esquema, como se muestra en la siguiente tabla.

```
create table tickit.public.test (c1) as select * from oldtable;
```

En este ejemplo, `tickit` es el nombre de base de datos y `public` es el nombre de esquema. Si la base de datos o el esquema no existen, la instrucción devuelve un error.

Si se proporciona un nombre de esquema, la tabla nueva se crea en ese esquema (suponiendo que el creador tiene acceso al esquema). El nombre de la tabla debe ser un nombre único para ese esquema. Si no se especifica un esquema, se crea la tabla a través del esquema de la base de datos actual. Si está creando una tabla temporal, no puede especificar un nombre de esquema dado que las tablas temporales existen en un esquema especial.

Pueden existir varias tablas temporales con el mismo nombre y al mismo tiempo en la misma base de datos si se crean en sesiones separadas. Estas tablas se asignan a esquemas diferentes.

column_name

El nombre de una columna en la tabla nueva. Si no se proporcionan nombres de columnas, se toman de los nombres de columnas de salida de la consulta. Se usan nombres de columnas predeterminados para las expresiones. Para obtener más información acerca de los nombres válidos, consulte Nombres e identificadores.

BACKUP { YES | NO }

Una cláusula que especifica si la tabla debe incluirse en instantáneas de clústeres manuales y automáticos.

Para las tablas, como las tablas provisionales que no contienen datos críticos, especifique BACKUP NO para ahorrar tiempo de procesamiento al momento de crear instantáneas y restaurar contenido a partir de ellas, y para reducir espacio de almacenamiento en Amazon Simple Storage Service. El ajuste BACKUP NO no tiene efecto en la replicación automática de datos a otros nodos dentro del clúster, por lo que las tablas con BACKUP NO especificado se restauran en el caso de un error del nodo. El valor predeterminado es BACKUP YES.

> **Note**
>
> Las tablas sin copia de seguridad no son compatibles con RA3 y sin servidor. Una tabla marcada como sin copia de seguridad en RA3 y sin servidor se tratará como una tabla permanente de la que siempre se hará una copia de seguridad al tomar una instantánea, y que se restaurará al restaurar desde una instantánea.

DISTSTYLE { AUTO | EVEN | KEY | ALL }

Se trata de la palabra clave que define el estilo de distribución de datos para toda la tabla. Amazon Redshift distribuye las filas de una tabla entre los nodos informáticos en función del estilo de distribución especificado en la tabla. El valor predeterminado es DISTSTYLE AUTO.

El estilo de distribución que selecciona para las tablas afecta el rendimiento global de la base de datos. Para obtener más información, consulte Distribución de datos para la optimización de consultas.

- AUTO: Amazon Redshift asigna un estilo de distribución óptimo en función de los datos de la tabla. Para ver el estilo de distribución aplicado a una tabla, consulte la tabla de catálogo del sistema PG_CLASS. Para obtener más información, consulte Visualización de los estilos de distribución.
- EVEN: Los datos de la tabla se distribuyen de manera uniforme en los nodos de un clúster en una distribución de turnos rotativos. Los ID de filas se utilizan para determinar la distribución, y se distribuye casi la misma cantidad de filas a cada nodo. Este es el método de distribución predeterminado.
- KEY: Los datos se distribuyen por los valores en la columna DISTKEY. Cuando establece las columnas de unión de las tablas de unión como claves de distribución, las filas de combinación de ambas tablas se colocan en los nodos de computación. Cuando se colocan los datos, el

optimizador puede realizar combinaciones de manera más eficiente. Si especifica DISTSTYLE KEY, debe asignar una columna DISTKEY.

- ALL: Una copia de toda la tabla se distribuye a cada nodo. El estilo de distribución garantiza que todas las filas obligatorias para cualquier combinación estén disponibles en todos los nodos, pero multiplica los requisitos de almacenamiento y aumenta el tiempo de carga y mantenimiento de la tabla. La distribución ALL puede mejorar el tiempo de ejecución cuando se usa con determinadas tablas de dimensión en las que la distribución KEY no es adecuada. No obstante, es conveniente analizar las mejoras de rendimiento en relación con los costos de mantenimiento.

DISTKEY (column)

Especifica un nombre de columna o número de posicionamiento para la clave de distribución. Use el nombre especificado en la lista de columnas opcional para la tabla o la lista de selección de la consulta. De manera opcional, use un número de posición, donde la primera columna seleccionada es 1, la segunda es 2, etc. Solo una columna de una tabla puede ser la clave de distribución:

- Si declara una columna como la columna DISTKEY, DISTSTYLE debe estar configurado en KEY o no estar configurado.
- Si no declara ninguna columna DISTKEY, puede establecer DISTSTYLE en EVEN.
- Si no especifica DISTKEY ni DISTSTYLE, CTAS determina el estilo de distribución para la nueva tabla en función del plan de consulta para la cláusula SELECT. Para obtener más información, consulte Herencia de atributos de columna y tabla.

Puede definir la misma columna como la clave de distribución y la clave de ordenación; este enfoque tiende a acelerar las combinaciones cuando la columna en cuestión es una columna de combinación en la consulta.

[COMPOUND | INTERLEAVED] SORTKEY (nombre_de_columna [, ... ])

Especifica una o más claves de ordenación para la tabla. Cuando carga datos en la tabla, los datos se ordenan por las columnas que se designan como claves de ordenación.

Puede especificar el estilo de ordenación COMPOUND o INTERLEAVED. La opción predeterminada es COMPOUND. Para obtener más información, consulte Claves de clasificación.

Puede definir un máximo de 400 columnas COMPOUND SORTKEY u 8 columnas INTERLEAVED SORTKEY por tabla.

Si no especifica SORTKEY, CTAS determina las claves de ordenación para la nueva tabla en función del plan de consulta para la cláusula SELECT. Para obtener más información, consulte Herencia de atributos de columna y tabla.

COMPOUND

Especifica que los datos se ordenan a través de una clave compuesta formada por todas las columnas enumeradas en el orden en que aparecen en la lista. Una clave de ordenación compuesta es útil cuando una consulta analiza filas según el orden de las columnas de ordenación. Los beneficios del rendimiento de ordenación con una clave compuesta se reducen cuando las consultas dependen de columnas de ordenación secundarias. Puede definir una cantidad máxima de 400 columnas COMPOUND SORTKEY por tabla.

INTERLEAVED

Especifica que los datos se ordenan a través de una clave de ordenación intercalada. Se puede especificar una cantidad máxima de ocho columnas para una clave de ordenación intercalada.

Una ordenación intercalada otorga el mismo peso a cada columna o subconjunto de columnas de la clave de ordenación. Por lo tanto, las consultas no dependen del orden que tengan las columnas en la clave de ordenación. Cuando una consulta usa una o más columnas de ordenación secundarias, la ordenación intercalada mejora considerablemente el rendimiento de la consulta. La ordenación intercalada tiene un pequeño costo general para las operaciones de carga y limpieza de datos.

AS query

Se trata de cualquier consulta (instrucción SELECT) que Amazon Redshift admita.

## Notas de uso de CTAS

### Límites

Amazon Redshift aplica una cuota del número de tablas por clúster según el tipo de nodo.

La cantidad máxima de caracteres para el nombre de una tabla es 127.

La cantidad máxima de columnas que se pueden definir en una única tabla es 1 600.

Herencia de atributos de columna y tabla

Las tablas CREATE TABLE AS (CTAS) no heredan restricciones, columnas de identidad, valores de columna predeterminados ni la clave principal de la tabla de la cual fueron creadas.

No se pueden especificar codificaciones de compresión de las columnas para tablas CTAS. Amazon Redshift asigna de forma automática la codificación de compresión de la siguiente manera:

- A las columnas que están definidas como claves de ordenación se les asigna una compresión RAW.
- A las columnas que están definidas como tipos de datos BOOLEAN, REAL, DOUBLE PRECISION, GEOMETRY o GEOGRAPHY se les asigna una compresión RAW.
- A las columnas que se definen como SMALLINT, INTEGER, BIGINT, DECIMAL, DATE, TIME, TIMETZ, TIMESTAMP o TIMESTAMPTZ se les asigna la compresión AZ64.
- A las columnas que se definen como CHAR, VARCHAR o VARBYTE se les asigna la compresión LZO.

Para obtener más información, consulte Codificaciones de compresión y Tipos de datos.

Para asignar explícitamente codificaciones de la columna, utilice CREATE TABLE.

CTAS determina el estilo de distribución y la clave de ordenación para la nueva tabla en función del plan de consulta para la cláusula SELECT.

Para consultas completas, como consultas que incluyen combinaciones, agregaciones, una cláusula de ordenación o una cláusula de límite, CTAS realiza su mejor esfuerzo por seleccionar el estilo de distribución y la clave de ordenación óptimos en función del plan de consulta.

> **Note**
>
> Para conseguir un mejor rendimiento con conjuntos de datos grandes o consultas complejas, recomendamos probar los conjuntos de datos típicos.

A menudo, puede predecir la clave de distribución y la clave de ordenación que CTAS selecciona al examinar el plan de consulta y ver cuáles son las columnas, si hay, que el optimizador de consultas selecciona para ordenar y distribuir datos. Si el nodo superior del plan de consulta es un análisis secuencial simple de una tabla única (XN Seq Scan), por lo general CTAS usa el estilo de distribución y la clave de ordenación de la tabla de origen. Si el nodo superior del plan de consulta no

es un análisis secuencial (como XN Limit, XN Sort, XN HashAggregate, etc.), CTAS realiza su mejor esfuerzo por seleccionar el estilo de distribución y la clave de ordenación óptimos en función del plan de consulta.

Por ejemplo, supongamos que crea cinco tablas con los siguientes tipos de cláusulas SELECT:

- Una instrucción de selección simple
- Una cláusula de límite
- Una cláusula de ordenación con LISTID
- Una cláusula de ordenación con QTYSOLD
- Una función de agregación SUM con una cláusula de agrupación.

En los siguientes ejemplos se muestra el plan de consulta para cada instrucción CTAS.

```
explain create table sales1_simple as select listid, dateid, qtysold from sales;
                           QUERY PLAN
-----------------------------------------------------------------
 XN Seq Scan on sales  (cost=0.00..1724.56 rows=172456 width=8)
(1 row)


explain create table sales2_limit as select listid, dateid, qtysold from sales limit
 100;
                              QUERY PLAN
----------------------------------------------------------------------
 XN Limit  (cost=0.00..1.00 rows=100 width=8)
   ->  XN Seq Scan on sales  (cost=0.00..1724.56 rows=172456 width=8)
(2 rows)


explain create table sales3_orderbylistid as select listid, dateid, qtysold from sales
 order by listid;
                               QUERY PLAN
-----------------------------------------------------------------------
 XN Sort  (cost=1000000016724.67..1000000017155.81 rows=172456 width=8)
   Sort Key: listid
   ->  XN Seq Scan on sales  (cost=0.00..1724.56 rows=172456 width=8)
(3 rows)
```

```
explain create table sales4_orderbyqty as select listid, dateid, qtysold from sales
 order by qtysold;
                              QUERY PLAN
-----------------------------------------------------------------------
 XN Sort  (cost=1000000016724.67..1000000017155.81 rows=172456 width=8)
   Sort Key: qtysold
   ->  XN Seq Scan on sales  (cost=0.00..1724.56 rows=172456 width=8)
(3 rows)


explain create table sales5_groupby as select listid, dateid, sum(qtysold) from sales
 group by listid, dateid;
                              QUERY PLAN
-----------------------------------------------------------------------
 XN HashAggregate  (cost=3017.98..3226.75 rows=83509 width=8)
   ->  XN Seq Scan on sales  (cost=0.00..1724.56 rows=172456 width=8)
(2 rows)
```

Para ver la clave de distribución y la clave de ordenación de cada tabla, consulte la tabla de catálogo del sistema PG_TABLE_DEF como se muestra a continuación.

```
select * from pg_table_def where tablename like 'sales%';

      tablename       |   column   | distkey | sortkey
----------------------+------------+---------+---------
 sales                | salesid    | f       |       0
 sales                | listid     | t       |       0
 sales                | sellerid   | f       |       0
 sales                | buyerid    | f       |       0
 sales                | eventid    | f       |       0
 sales                | dateid     | f       |       1
 sales                | qtysold    | f       |       0
 sales                | pricepaid  | f       |       0
 sales                | commission | f       |       0
 sales                | saletime   | f       |       0
 sales1_simple        | listid     | t       |       0
 sales1_simple        | dateid     | f       |       1
 sales1_simple        | qtysold    | f       |       0
 sales2_limit         | listid     | f       |       0
 sales2_limit         | dateid     | f       |       0
 sales2_limit         | qtysold    | f       |       0
 sales3_orderbylistid | listid     | t       |       1
 sales3_orderbylistid | dateid     | f       |       0
```

```
sales3_orderbylistid | qtysold    | f       |       0
sales4_orderbyqty    | listid     | t       |       0
sales4_orderbyqty    | dateid     | f       |       0
sales4_orderbyqty    | qtysold    | f       |       1
sales5_groupby       | listid     | f       |       0
sales5_groupby       | dateid     | f       |       0
sales5_groupby       | sum        | f       |       0
```

En la siguiente tabla se resumen los resultados. Para simplificar, omitimos detalles de ancho, filas y costo del plan de explicación.

| Tabla | Instrucción CTAS SELECT | Explicar nodo superior del plan | Clave de distribución | Clave de clasificación |
|---|---|---|---|---|
| S1_SIMPLE | `select listid, dateid, qtysold from sales` | `XN Seq Scan on sales ...` | LISTID | DATEID |
| S2_LIMIT | `select listid, dateid, qtysold from sales limit 100` | `XN Limit ...` | Ninguno (EVEN) | Ninguno |
| S3_ORDER_BY_LISTID | `select listid, dateid, qtysold from sales order by listid` | `XN Sort ...`<br>`Sort Key: listid` | LISTID | LISTID |
| S4_ORDER_BY_QTY | `select listid, dateid, qtysold from sales order by qtysold` | `XN Sort ...`<br>`Sort Key: qtysold` | LISTID | QTYSOLD |
| S5_GROUP_BY | `select listid, dateid, sum(qtysold) from sales group by listid, dateid` | `XN HashAggregate ...` | Ninguno (EVEN) | Ninguno |

Puede especificar explícitamente el estilo de distribución y la clave de ordenación en la instrucción de CTAS. Por ejemplo, la siguiente instrucción crea una tabla con la distribución EVEN y especifica SALESID como la clave de ordenación.

```
create table sales_disteven
diststyle even
sortkey (salesid)
as
select eventid, venueid, dateid, eventname
from event;
```

Codificación de compresión

ENCODE AUTO se usa como la opción predeterminada para las tablas. Amazon Redshift administra de forma automática la codificación de compresión para todas las columnas de la tabla.

Distribución de datos entrantes

Cuando el esquema de distribución hash de los datos entrantes coincide con el de la tabla de destino, no es necesaria la distribución física de datos cuando se cargan los datos. Por ejemplo, si se configura una clave de distribución para la nueva tabla y los datos se insertan desde otra tabla que está distribuida en la misma columna clave, los datos se cargan en el lugar a través de los mismos nodos y sectores. No obstante, si las tablas de origen y destino se configuran en distribución EVEN, los datos se redistribuyen en la tabla de destino.

Operaciones ANALYZE automáticas

Amazon Redshift analiza de forma automática las tablas que usted crea con los comandos CTAS. No necesita ejecutar el comando ANALYZE en estas tablas cuando se crean. Si las modifica, debe analizarlas de la misma manera en que analiza las demás tablas.

## Ejemplos de CTAS

En el siguiente ejemplo, se crea una tabla denominada EVENT_BACKUP para la tabla EVENT:

```
create table event_backup as select * from event;
```

La tabla resultante hereda las claves de distribución y ordenación de la tabla EVENT.

```
select "column", type, encoding, distkey, sortkey
from pg_table_def where tablename = 'event_backup';
```

```
column    | type                        | encoding | distkey | sortkey
----------+-----------------------------+----------+---------+--------
catid     | smallint                    | none     | false   |       0
dateid    | smallint                    | none     | false   |       1
eventid   | integer                     | none     | true    |       0
eventname | character varying(200)      | none     | false   |       0
starttime | timestamp without time zone | none     | false   |       0
venueid   | smallint                    | none     | false   |       0
```

El siguiente comando crea una tabla nueva denominada EVENTDISTSORT al seleccionar cuatro columnas de la tabla EVENT. La tabla nueva es distribuida por EVENTID y ordenada por EVENTID y DATEID:

```
create table eventdistsort
distkey (1)
sortkey (1,3)
as
select eventid, venueid, dateid, eventname
from event;
```

El resultado es el siguiente.

```
select "column", type, encoding, distkey, sortkey
from pg_table_def where tablename = 'eventdistsort';

column   |          type          | encoding | distkey | sortkey
---------+------------------------+----------+---------+-------
eventid   | integer               | none     | t       | 1
venueid   | smallint              | none     | f       | 0
dateid    | smallint              | none     | f       | 2
eventname | character varying(200)| none     | f       | 0
```

Podría crear exactamente la misma tabla al utilizar nombres de columnas para las claves de distribución y ordenación. Por ejemplo:

```
create table eventdistsort1
distkey (eventid)
sortkey (eventid, dateid)
as
select eventid, venueid, dateid, eventname
```

```
from event;
```

La siguiente instrucción aplica una distribución uniforme a la tabla, pero no define una clave de ordenación explícita:

```
create table eventdisteven
diststyle even
as
select eventid, venueid, dateid, eventname
from event;
```

La tabla no hereda la clave de ordenación de la tabla EVENT (EVENTID) porque la distribución EVEN está especificada para la tabla nueva. La tabla nueva no tiene clave de ordenación ni clave de distribución.

```
select "column", type, encoding, distkey, sortkey
from pg_table_def where tablename = 'eventdisteven';

column    |          type          | encoding | distkey | sortkey
----------+------------------------+----------+---------+---------
eventid   | integer                | none     | f       | 0
venueid   | smallint               | none     | f       | 0
dateid    | smallint               | none     | f       | 0
eventname | character varying(200) | none     | f       | 0
```

La siguiente instrucción aplica distribución uniforme y define una clave de ordenación:

```
create table eventdistevensort diststyle even sortkey (venueid)
as select eventid, venueid, dateid, eventname from event;
```

La tabla resultante tiene una clave de ordenación pero no tiene una clave de distribución.

```
select "column", type, encoding, distkey, sortkey
from pg_table_def where tablename = 'eventdistevensort';

column    |          type          | encoding | distkey | sortkey
----------+------------------------+----------+---------+-------
eventid   | integer                | none     | f       | 0
venueid   | smallint               | none     | f       | 1
dateid    | smallint               | none     | f       | 0
eventname | character varying(200) | none     | f       | 0
```

La siguiente instrucción redistribuye la tabla EVENT en una columna de clave diferente de los datos entrantes, que está ordenada en la columna EVENTID, y no define una columna SORTKEY. Por lo tanto, la tabla no se ordena.

```
create table venuedistevent distkey(venueid)
as select * from event;
```

El resultado es el siguiente.

```
select "column", type, encoding, distkey, sortkey
from pg_table_def where tablename = 'venuedistevent';

 column   |            type             | encoding | distkey | sortkey
----------+-----------------------------+----------+---------+-------
eventid   | integer                     | none     | f       | 0
venueid   | smallint                    | none     | t       | 0
catid     | smallint                    | none     | f       | 0
dateid    | smallint                    | none     | f       | 0
eventname | character varying(200)      | none     | f       | 0
starttime | timestamp without time zone | none     | f       | 0
```

# CREAR USUARIO

Crea un nuevo usuario de base de datos. Los usuarios de base de datos pueden recuperar datos, ejecutar comandos y realizar otras acciones en una base de datos, en función de sus privilegios y roles. Debe ser un superusuario de base de datos para ejecutar este comando.

## Privilegios necesarios

Los siguientes privilegios son necesarios para CREATE USER:

- Superusuario
- Usuarios con el privilegio CREATE USER

## Sintaxis

```
CREATE USER name [ WITH ]
PASSWORD { 'password' | 'md5hash' | 'sha256hash' | DISABLE }
[ option [ ... ] ]
```

```
where option can be:

CREATEDB | NOCREATEDB
| CREATEUSER | NOCREATEUSER
| SYSLOG ACCESS { RESTRICTED | UNRESTRICTED }
| IN GROUP groupname [, ... ]
| VALID UNTIL 'abstime'
| CONNECTION LIMIT { limit | UNLIMITED }
| SESSION TIMEOUT limit
| EXTERNALID external_id
```

## Parámetros

name

Nombre del usuario que se va a crear. El nombre de usuario no puede ser PUBLIC. Para obtener más información acerca de los nombres válidos, consulte Nombres e identificadores.

WITH

Palabra clave opcional. Amazon Redshift ignora la palabra WITH.

PASSWORD { 'password' | 'md5hash' | 'sha256hash' | DISABLE }

Establece la contraseña del usuario.

De manera predeterminada, los usuarios pueden cambiar sus propias contraseñas, a menos que la contraseña esté deshabilitada. Para deshabilitar la contraseña de un usuario, especifique DISABLE. Cuando se deshabilita la contraseña de un usuario, se elimina del sistema y el usuario solo puede iniciar sesión con credenciales de usuario de AWS Identity and Access Management IAM temporales. Para obtener más información, consulte Uso de la autenticación de IAM para generar credenciales de usuario de base de datos. Solo un superusuario puede habilitar o deshabilitar contraseñas. No puede deshabilitar la contraseña de un superusuario. Para habilitar una contraseña, ejecute ALTER USER y especifique una contraseña.

Puede especificar la contraseña en texto sin cifrar, como una cadena hash MD5 o como una cadena de algoritmo hash seguro-256.

> **Note**
>
> Cuando lanza un clúster nuevo a través de la AWS Management Console, la AWS CLI o la API de Amazon Redshift, debe proporcionar una contraseña de texto sin cifrar para

el usuario inicial de la base de datos. Puede cambiar la contraseña en otro momento al utilizar ALTER USER.

Para texto sin cifrar, la contraseña debe cumplir con las siguientes restricciones:

- Tener entre 8 y 64 caracteres de longitud.
- Contener al menos una letra en mayúsculas, una letra en minúsculas y un número.
- Puede utilizar cualquier carácter ASCII con códigos ASCII del rango 33-126, excepto ' (comilla simple), " (comilla doble), \, / o @.

Como una alternativa más segura para especificar el parámetro de contraseña CREATE USER como texto sin cifrar, puede especificar un hash MD5 de una cadena que incluya la contraseña y el nombre de usuario.

> **Note**
>
> Cuando especifica una cadena hash MD5, el comando CREATE USER busca una cadena hash MD5 válida, pero no valida la porción de la contraseña de la cadena. En este caso es posible crear una contraseña, como una cadena vacía, que no puede usar para ingresar en la base de datos.

Siga estos pasos para especificar una contraseña MD5:

1. Concatene la contraseña y el nombre de usuario.

   Por ejemplo, para la contraseña `ez` y el usuario `user1`, la cadena concatenada es `ezuser1`.

2. Convierta la cadena concatenada en una cadena hash MD5 de 32 caracteres. Puede usar cualquier utilidad MD5 para crear la cadena hash. En el siguiente ejemplo, se utiliza Función MD5 de Amazon Redshift y el operador de concatenación (||) para devolver una cadena hash MD5 de 32 caracteres.

```
select md5('ez' || 'user1');

md5
----------------------------------
153c434b4b77c89e6b94f12c5393af5b
```

3. Concatene "md5" delante de la cadena hash MD5 y proporcione la cadena concatenada como argumento md5hash.

```
create user user1 password 'md5153c434b4b77c89e6b94f12c5393af5b';
```

4. Inicie sesión en la base de datos con las credenciales de inicio de sesión.

   Para este ejemplo, ingrese como `user1` con la contraseña `ez`.

Otra alternativa segura es especificar un hash de algoritmo hash seguro-256 de una cadena de contraseñas; o puede proporcionar su propio resumen de algoritmo hash seguro-256 válido y la sal de 256 bits que se utilizó para crear el resumen.

- Resumen: salida de una función de hash.
- Sal: datos generados aleatoriamente que se combinan con la contraseña para ayudar a reducir los patrones en la salida de la función hash.

```
'sha256|Mypassword'
```

```
'sha256|digest|256-bit-salt'
```

En el siguiente ejemplo, Amazon Redshift genera y administra la sal.

```
CREATE USER admin PASSWORD 'sha256|Mypassword1';
```

En el siguiente ejemplo, se proporciona un resumen de algoritmo hash seguro-256 válido y una sal de 256 bits que se utilizó para crear el resumen.

Para especificar una contraseña y cifrarla con tu propio valor de salt, sigue estos pasos:

1. Cree un valor de salt de 256 bits. Puede obtener un valor de salt utilizando cualquier generador de cadenas hexadecimales para generar una cadena de 64 caracteres de longitud. En este ejemplo, el valor de salt es `c721bff5d9042cf541ff7b9d48fa8a6e545c19a763e3710151f9513038b0f6c6`.
2. Utilice la función FROM_HEX para convertir su valor de salt en binario. Esto se debe a que la función SHA2 requiere la representación binaria del valor de salt. Utilice la siguiente instrucción.

```
SELECT
 FROM_HEX('c721bff5d9042cf541ff7b9d48fa8a6e545c19a763e3710151f9513038b0f6c6');
```

3. Utilice la función CONCAT para anexar su valor de salt a su contraseña. Para este ejemplo, la contraseña es `Mypassword1`. Utilice la siguiente instrucción.

```
SELECT
 CONCAT('Mypassword1',FROM_HEX('c721bff5d9042cf541ff7b9d48fa8a6e545c19a763e3710151f951303
```

4. Utilice la función SHA2 para crear un resumen a partir de la combinación de contraseña y valor de salt. Utilice la siguiente instrucción.

```
SELECT
 SHA2(CONCAT('Mypassword1',FROM_HEX('c721bff5d9042cf541ff7b9d48fa8a6e545c19a763e3710151f9
 0);
```

5. Mediante el resumen y el valor de sal de los pasos anteriores, cree el usuario. Utilice la siguiente instrucción.

```
CREATE USER admin PASSWORD 'sha256|
821708135fcc42eb3afda85286dee0ed15c2c461d000291609f77eb113073ec2|
c721bff5d9042cf541ff7b9d48fa8a6e545c19a763e3710151f9513038b0f6c6';
```

6. Inicie sesión en la base de datos con las credenciales de inicio de sesión.

   Para este ejemplo, ingrese como `admin` con la contraseña `Mypassword1`.

Si configura una contraseña en texto sin formato sin especificar la función de hash, se genera un resumen MD5 utilizando el nombre de usuario como la sal.

CREATEDB | NOCREATEDB

La opción CREATEDB permite que el usuario cree bases de datos. El valor predeterminado es NOCREATEDB.

CREATEUSER | NOCREATEUSER

La opción CREATEUSER crea un superusuario con todos los privilegios de base de datos, incluido CREATE USER. El valor predeterminado es NOCREATEUSER. Para obtener más información, consulte superuser.

SYSLOG ACCESS { RESTRICTED | UNRESTRICTED }

Se trata de una cláusula que especifica el nivel de acceso a las vistas y las tablas de sistema de Amazon Redshift que tiene el usuario.

Los usuarios normales que tengan el permiso SYSLOG ACCESS RESTRICTED solo pueden ver las filas generadas por ese usuario en las tablas y vistas del sistema visibles para el usuario. El valor predeterminado es RESTRICTED.

Los usuarios normales que tengan el permiso SYSLOG ACCESS UNRESTRICTED pueden ver todas las filas en las tablas y vistas del sistema visibles para el usuario, incluidas las filas generadas por otro usuario. UNRESTRICTED no permite que los usuarios normales puedan obtener acceso a las tablas visibles para los superusuarios. Solo los superusuarios pueden ver estas tablas.

> **Note**
>
> Al brindar a un usuario acceso sin restricciones a las tablas del sistema, le proporciona la visibilidad necesaria para ver los datos generados por otros usuarios. Por ejemplo, STL_QUERY y STL_QUERYTEXT contienen todo el texto de las instrucciones INSERT, UPDATE y DELETE, que podrían incluir datos confidenciales generados por los usuarios.

Todas las filas de SVV_TRANSACTIONS son visibles para todos los usuarios.

Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

IN GROUP groupname

Especifica el nombre de un grupo existente al que pertenece el usuario. Se pueden enumerar varios nombres de grupos.

VALID UNTIL abstime

La opción VALID UNTIL establece un tiempo absoluto después del cual la contraseña de la cuenta de usuario ya no es válida. Por defecto, la contraseña no tiene límite de tiempo.

CONNECTION LIMIT { limit | UNLIMITED }

La cantidad máxima de conexiones a la base de datos que el usuario puede tener abiertas al mismo tiempo. Este límite no se aplica a los superusuarios. Use la palabra clave UNLIMITED

para permitir la cantidad máxima de conexiones simultáneas. También puede aplicar un límite en la cantidad de conexiones de cada base de datos. Para obtener más información, consulte CREATE DATABASE. El valor predeterminado es UNLIMITED. Para ver las conexiones actuales, consulte la vista del sistema STV_SESSIONS.

> **Note**
>
> Si se aplican los límites de conexión tanto para usuarios como para bases de datos, debe haber una ranura de conexión sin utilizar disponible dentro de ambos límites cuando un usuario intenta conectarse.

SESSION TIMEOUT limit

El tiempo máximo en segundos durante el cual una sesión permanece inactiva o en reposo. El rango es de 60 segundos (un minuto) a 1 728 000 segundos (20 días). Si no se establece el tiempo de espera de la sesión para el usuario, se aplica el ajuste del clúster. Para obtener más información, consulte Cuotas y límites de Amazon Redshift en la Guía de administración de Amazon Redshift.

Cuando se establece el tiempo de espera de la sesión, se aplica solo a las sesiones nuevas.

Para ver información sobre las sesiones de usuario activas, incluidos la hora de inicio, el nombre de usuario y el tiempo de espera de la sesión, consulte la vista de sistema STV_SESSIONS. Para ver información sobre el historial de sesiones del usuario, consulte la vista STL_SESSIONS. Para recuperar información acerca de los usuarios de las bases de datos, incluidos los valores de tiempo de espera de la sesión, consulte la vista SVL_USER_INFO.

EXTERNALID external_id

El identificador del usuario, que está asociado a un proveedor de identidades. El usuario debe tener desactivada la contraseña. Para obtener más información, consulte Federación de proveedores de identidades (IdP) nativos para Amazon Redshift.

Notas de uso

Por defecto, todos los usuarios tienen privilegios CREATE y USAGE en el esquema PUBLIC. Para no permitirles a los usuarios crear objetos en el esquema PUBLIC de una base de datos, use el comando REVOKE para eliminar ese privilegio.

Cuando use la autenticación de IAM para crear credenciales de usuario de la base de datos, tal vez le convenga crear un superusuario que solo pueda iniciar sesión con credenciales temporales. Puede deshabilitar la contraseña de un superusuario, pero no puede crear una contraseña desconocida mediante una cadena hash MD5 generada aleatoriamente.

```
create user iam_superuser password 'md5A1234567890123456780123456789012' createuser;
```

El caso de un nombre de usuario entre comillas dobles se conserva siempre, independientemente del valor de la opción de configuración `enable_case_sensitive_identifier`. Para obtener más información, consulte enable_case_sensitive_identifier.

## Ejemplos

El siguiente comando crea un usuario denominado dbuser con la contraseña “abcD1234”, privilegios de creación de base de datos y un límite de conexión de 30.

```
create user dbuser with password 'abcD1234' createdb connection limit 30;
```

Consulte la tabla de catálogos PG_USER_INFO para ver detalles de un usuario de la base de datos.

```
select * from pg_user_info;

 usename   | usesysid | usecreatedb | usesuper | usecatupd | passwd   | valuntil |
 useconfig | useconnlimit
-----------+----------+-------------+----------+-----------+----------+----------
+-----------+-------------
 rdsdb     |        1 | true        | true     | true      | ******** | infinity |
      |
 adminuser |      100 | true        | true     | false     | ******** |          |
      | UNLIMITED
 dbuser    |      102 | true        | false    | false     | ******** |          |
      | 30
```

En el siguiente ejemplo, la contraseña de la cuenta es válida hasta el 10 de junio de 2017.

```
create user dbuser with password 'abcD1234' valid until '2017-06-10';
```

En el siguiente ejemplo, se crea un usuario con una contraseña que distingue entre mayúsculas y minúsculas que contiene caracteres especiales.

```
create user newman with password '@AbC4321!';
```

Para usar una barra oblicua inversa ("\") en la contraseña MD5, encierre la barra oblicua inversa con una barra oblicua inversa en la cadena de origen. En el siguiente ejemplo, se crea un usuario denominado `slashpass` con una barra oblicua inversa ("\") como la contraseña.

```
select md5('\\'||'slashpass');

md5
--------------------------------
0c983d1a624280812631c5389e60d48c
```

Cree un usuario con la contraseña md5.

```
create user slashpass password 'md50c983d1a624280812631c5389e60d48c';
```

En el siguiente ejemplo, se crea un usuario denominado `dbuser` con un tiempo de espera para las sesiones inactivas establecido en 120 segundos.

```
CREATE USER dbuser password 'abcD1234' SESSION TIMEOUT 120;
```

El siguiente ejemplo crea una usuario denominado `bob`. El espacio de nombres es `myco_aad`. Solo es un ejemplo. Para ejecutar el comando correctamente, debe tener un proveedor de identidades registrado. Para obtener más información, consulte Federación de proveedores de identidades nativos (IdP) para Amazon Redshift.

```
CREATE USER myco_aad:bob EXTERNALID "ABC123" PASSWORD DISABLE;
```

# CREATE VIEW

Crea una vista en una base de datos. La vista no está materializada físicamente; la consulta que define la vista se ejecuta cada vez que se hace referencia a dicha vista en una consulta. Para crear una vista con una tabla externa, incluya la cláusula WITH NO SCHEMA BINDING.

Para crear una vista estándar, debe obtener acceso a las tablas o vistas subyacentes. Para consultar una vista estándar, debe seleccionar los premisos de la vista, pero no es necesario seleccionar permisos para las tablas subyacentes. Si crea una vista que haga referencia a una tabla o vista de otro esquema o si crea una vista que haga referencia a una vista materializada, es posible que

necesite permisos de uso. Para consultar una vista de enlace tardío, debe seleccionar los permisos de la propia vista de enlace tardío. También debe asegurarse de que el propietario de la vista de enlace tardío tiene permisos de selección para los objetos de referencia (tablas, vistas o funciones definidas por el usuario). Para obtener más información acerca de las vistas de enlace tardío, consulte Notas de uso.

## Permisos necesarios

Para utilizar CREATE VIEW, se requiere uno de los siguientes permisos.

- Para crear una vista con CREATE [ OR REPLACE ] VIEW:
  - Superusuario
  - Usuarios con el permiso CREATE [ REPLACE ] VIEW
- Para reemplazar una vista existente mediante CREATE OR REPLACE VIEW:
  - Superusuario
  - Usuarios con el permiso CREATE [ OR REPLACE ] VIEW
  - Propietario de la vista

Si un usuario desea acceder a una vista que incorpora una función definida por el usuario, debe contar con el permiso EXECUTE en esa función.

## Sintaxis

```
CREATE [ OR REPLACE ] VIEW name [ ( column_name [, ...] ) ] AS query
[ WITH NO SCHEMA BINDING ]
```

## Parámetros

OR REPLACE

Si ya existe una vista con el mismo nombre, se sustituye la vista. Solo puede reemplazar una vista con una nueva consulta que genere el conjunto idéntico de columnas, con los mismos nombres de columnas y los mismos tipos de datos. CREATE OR REPLACE VIEW bloquea la vista para lecturas y escrituras hasta que se completa la operación.

Cuando se reemplaza una vista, se conservan sus otras propiedades, como la propiedad y los privilegios concedidos.

name

El nombre de la vista. Si se proporciona un nombre de esquema (como `myschema.myview`), la vista se crea con el esquema especificado. De lo contrario, la vista se crea en el esquema actual. El nombre de la vista debe ser distinto al de cualquier otra vista o tabla en el mismo esquema.

Si especifica un nombre de vista que comienza con "**#**", la vista se crea como una vista temporal que solo es visible en la sesión actual.

Para obtener más información acerca de los nombres válidos, consulte Nombres e identificadores. No puede crear tablas o vistas en las bases de datos del sistema template0, template1, padb_harvest o sys:internal.

column_name

Lista opcional de nombres que se utilizarán para las columnas en la vista. Si no se proporcionan nombres de columnas, se derivan de la consulta. La cantidad máxima de columnas que se pueden definir en una única vista es 1 600.

consulta

Una consulta (en la forma de una instrucción SELECT) que se evalúa en una tabla. Esta tabla define las columnas y filas de la vista.

WITH NO SCHEMA BINDING

Cláusula que especifica que la vista no está vinculada a los objetos de base de datos subyacentes, como tablas y funciones definidas por el usuario. Por tanto, no existe ninguna dependencia entre la vista y los objetos a los que hace referencia. Puede crear una vista aunque no existan los objetos a los que se hace referencia. Como no hay dependencias, puede quitar o modificar un objeto al que se referencie sin que esto afecte a la vista. Amazon Redshift no verifica si existen dependencias hasta que se consulta la vista. Para ver los detalles de las vistas de enlace tardío, ejecute la función PG_GET_LATE_BINDING_VIEW_COLS.

Cuando incluya la cláusula WITH NO SCHEMA BINDING, las tablas y vistas a las que se hace referencia en la instrucción SELECT deben incluir el nombre de un esquema. El esquema debe existir cuando se crea la vista, aunque no exista la tabla a la que se hace referencia. Por ejemplo, la siguiente instrucción devuelve un error.

```
create view myevent as select eventname from event
with no schema binding;
```

La siguiente instrucción se ejecuta correctamente.

```
create view myevent as select eventname from public.event
with no schema binding;
```

> **Note**
>
> No se puede actualizar, insertar en o eliminar de una vista.

## Notas de uso

### Vistas de enlace en tiempo de ejecución

Una vista de enlace en tiempo de ejecución comprueba los objetos de base de datos subyacentes, como tablas y otras vistas, hasta que se consulta la vista. Por consiguiente, puede modificar o borrar los objetos subyacentes sin borrar o volver a crear la vista. Si borra objetos subyacentes, las consultas a la vista de enlace de tiempo de ejecución darán error. Si la consulta de la vista de enlace de tiempo de ejecución hace referencia a columnas del objeto subyacente que no están presentes, la consulta no se ejecutará correctamente.

Si borra y luego vuelve a crear una tabla o vista subyacente de la vista de enlace de tiempo de ejecución, el nuevo objeto se creará con los permisos de acceso predeterminados. Es posible que deba conceder permisos para los objetos subyacentes a los usuarios que consultarán la vista.

Para crear una vista de enlace en tiempo de ejecución, incluya la cláusula WITH NO SCHEMA BINDING. El siguiente ejemplo crea un vista sin enlace con el esquema.

```
create view event_vw as select * from public.event
with no schema binding;
```

```
select * from event_vw limit 1;

eventid | venueid | catid | dateid | eventname     | starttime
--------+---------+-------+--------+---------------+--------------------
      2 |     306 |     8 |   2114 | Boris Godunov | 2008-10-15 20:00:00
```

El siguiente ejemplo muestra que puede modificar o borrar una tabla subyacente sin volver a crear la vista.

```
alter table event rename column eventname to title;
```

```
select * from event_vw limit 1;

eventid | venueid | catid | dateid | title         | starttime
--------+---------+-------+--------+---------------+--------------------
      2 |     306 |     8 |   2114 | Boris Godunov | 2008-10-15 20:00:00
```

Solo puede referenciar las tablas externas de Amazon Redshift Spectrum en una vista de enlace de tiempo de ejecución. Una aplicación de las vistas de enlace de tiempo de ejecución es consultar las tablas de Amazon Redshift y de Redshift Spectrum. Por ejemplo, puede utilizar el comando UNLOAD para archivar datos más antiguos en Amazon S3. Luego, cree una tabla externa de Redshift Spectrum que referencie los datos de Amazon S3 y cree una vista que consulte ambas tablas. En el siguiente ejemplo, se utiliza una cláusula UNION ALL para combinar la tabla SALES de Amazon Redshift y la tabla SPECTRUM.SALES de Redshift Spectrum.

```
create view sales_vw as
select * from public.sales
union all
select * from spectrum.sales
with no schema binding;
```

Para obtener más información acerca de la creación de tablas externas de Redshift Spectrum, incluida la tabla SPECTRUM.SALES, consulte Introducción a Amazon Redshift Spectrum.

> **Important**
>
> Al crear una vista estándar a partir de una vista de enlace tardío, la definición estándar de la vista contiene la definición de la vista de enlace tardío en el momento en que se creó la vista estándar, incluido el propietario de la vista de enlace tardío. Si realiza un cambio en la vista de enlace tardío subyacente, esos cambios no se utilizarán en la vista estándar hasta que esta se vuelva a crear. Por lo tanto, cuando se consulta la vista estándar, siempre utilizará la definición de la vista de enlace tardío y el propietario de la vista de enlace tardío para comprobar los permisos en el momento de crear la vista estándar.

Para actualizar la vista estándar y hacer referencia a la definición más reciente de la vista de enlace tardío, ejecute CREATE OR REPLACE VIEW con la definición de vista inicial que ha utilizado para crear la vista estándar.

Consulte el siguiente ejemplo de creación de una vista estándar a partir de una vista de enlace tardío.

```
create view sales_vw_lbv as
select * from public.sales
with no schema binding;

show view sales_vw_lbv;
                            Show View DDL statement
-------------------------------------------------------------------------------
 create view sales_vw_lbv as select * from public.sales with no schema binding;
(1 row)

create view sales_vw as
select * from sales_vw_lbv;

show view sales_vw;
                                                 Show View DDL statement
-------------------------------------------------------------------------------------------------------
 SELECT sales_vw_lbv.price, sales_vw_lbv."region" FROM (SELECT sales.price,
 sales."region" FROM sales) sales_vw_lbv;
(1 row)
```

Tenga en cuenta que la vista de enlace tardío, tal como se muestra en la instrucción DDL de la vista estándar, se define al crear la vista estándar y no se actualizará con los cambios que realice posteriormente en la vista de enlace tardío.

## Ejemplos

Los comandos de ejemplo utilizan un conjunto de objetos y datos de muestra denominado base de datos TICKIT. Para obtener más información, consulte Base de datos de muestra.

El siguiente comando crea una vista denominada myevent a partir de una tabla llamada EVENT.

```
create view myevent as select eventname from event
where eventname = 'LeAnn Rimes';
```

El siguiente comando crea una vista denominada myuser a partir de una tabla llamada USERS.

```
create view myuser as select lastname from users;
```

El siguiente comando crea o sustituye una vista denominada myuser a partir de una tabla llamada USERS.

```
create or replace view myuser as select lastname from users;
```

El siguiente ejemplo crea un vista sin enlace con el esquema.

```
create view myevent as select eventname from public.event
with no schema binding;
```

# DEALLOCATE

Anula la asignación de una instrucción preparada.

## Sintaxis

```
DEALLOCATE [PREPARE] plan_name
```

## Parámetros

PREPARE

Esta palabra clave es opcional y se ignora.

plan_name

El nombre de la instrucción preparada para la que desea cancelar la asignación.

## Notas de uso

DEALLOCATE se usa para cancelar la asignación de una instrucción SQL preparada con anterioridad. Si no desasigna explícitamente una instrucción preparada, se desasignará cuando finalice la sesión actual. Para obtener más información acerca de instrucciones preparadas, consulte PREPARE.

## Véase también

EXECUTE, PREPARE

# DECLARE

Define un nuevo cursor. Utilice un cursor para recuperar algunas filas a la vez del conjunto de resultados de una consulta más grande.

Cuando se recupera la primera fila de un cursor, el conjunto completo de resultados se materializa en el nodo principal, en la memoria o en el disco, si es necesario. Debido al posible impacto negativo en el rendimiento del uso de cursores con conjuntos de resultados grandes, recomendamos usar enfoques alternativos cuando sea posible. Para obtener más información, consulte Consideraciones acerca del rendimiento cuando se utilizan cursores.

Debe declarar un cursor dentro de un bloque de transacción. Solo se puede abrir un cursor a la vez por sesión.

Para obtener más información, consulte FETCH, CLOSE.

## Sintaxis

```
DECLARE cursor_name CURSOR FOR query
```

## Parámetros

cursor_name

Nombre del nuevo cursor.

consulta

Una instrucción SELECT que rellena el cursor.

## Notas de uso acerca de DECLARE CURSOR

Si su aplicación cliente utiliza una conexión Open Database Connectivity (ODBC, Conectividad de base de datos abierta) y su consulta crea un conjunto de resultados que es demasiado grande para entrar en la memoria, puede transmitir el conjunto de resultados a su aplicación cliente por medio del cursor. Cuando utiliza un cursor, el conjunto completo de resultados se materializa en el nodo principal y, luego, el cliente puede recuperar los resultados gradualmente.

> **Note**
>
> Para habilitar cursores en ODBC para Microsoft Windows, habilite la opción Use Declare/Fetch (Utilizar Declarar/Recuperar) en el nombre de origen de datos (DSN) de ODBC que utiliza para Amazon Redshift. En los clústeres de varios nodos, es recomendable establecer el tamaño de la caché de ODBC en 4000 o un valor superior utilizando el campo Cache Size (Tamaño de caché) del cuadro de diálogo de opciones del DSN de ODBC para minimizar los recorridos de ida y vuelta. En un clúster de un nodo, configure el tamaño de la caché en 1 000.

Debido al posible impacto negativo en el rendimiento del uso de cursores, recomendamos usar enfoques alternativos cuando sea posible. Para obtener más información, consulte Consideraciones acerca del rendimiento cuando se utilizan cursores.

Los cursores de Amazon Redshift se admiten con las siguientes limitaciones:

- Solo se puede abrir un cursor a la vez por sesión.
- Los cursores deben utilizarse dentro de una transacción (BEGIN … END).
- El tamaño máximo del conjunto de resultados acumulables para todos los cursores está limitado en función del tipo de nodo del clúster. Si necesita conjuntos de resultados más grandes, puede cambiar el tamaño a una configuración de nodos XL o 8XL.

  Para obtener más información, consulte Restricciones del cursor.

## Restricciones del cursor

Cuando se recupera la primera fila de un cursor, el conjunto completo de resultados se materializa en el nodo principal. Si el conjunto de resultados no entra en la memoria, se escribe en el disco, según sea necesario. Para proteger la integridad del nodo principal, Amazon Redshift aplica restricciones al tamaño de todos los conjuntos de resultados del cursor, en función del tipo de nodo del clúster.

En la tabla siguiente se muestra el tamaño máximo total del conjunto de resultados para cada tipo de nodo de clúster. Los tamaños máximos del conjunto de resultados se expresan en megabytes.

| Tipo de nodo | Conjunto de resultados máximo por clúster (MB) |
|---|---|
| Varios nodos DC2 Large | 192 000 |

| Tipo de nodo | Conjunto de resultados máximo por clúster (MB) |
|---|---|
| Un solo nodo DC2 Large | 8000 |
| Varios nodos DC2 8XL | 3 200 000 |
| Varios nodos RA3 16XL | 14 400 000 |
| Varios nodos RA3 4XL | 3 200 000 |
| Varios nodos RA3 XLPLUS | 1 000 000 |
| Nodo único RA3 XLPLUS | 64 000 |
| Varios nodos RA3 LARGE | 240 000 |
| Un nodo RA3 LARGE | 8000 |
| Amazon Redshift sin servidor | 150 000 |

Para ver la configuración del cursor activo para un clúster, consulte la tabla del sistema STV_CURSOR_CONFIGURATION como superusuario. Para ver el estado de los cursores activos, consulte la table del sistema STV_ACTIVE_CURSORS. El usuario puede ver solo las filas de los propios cursores de un usuario, pero un superusuario puede ver todos los cursores.

## Consideraciones acerca del rendimiento cuando se utilizan cursores

Debido a que los cursores materializan el conjunto completo de resultados en el nodo principal antes de comenzar a devolver resultados al cliente, el uso de cursores con conjuntos de resultados grandes puede tener un impacto negativo en el rendimiento. No recomendamos utilizar cursores con conjuntos de resultados muy grandes. En algunos casos, como cuando la aplicación usa una conexión ODBC, los cursores pueden ser la única solución viable. Si es posible, recomendamos utilizar estas alternativas:

- Utilice UNLOAD para exportar una tabla grande. Cuando utiliza UNLOAD, los nodos informáticos trabajan en paralelo para transferir los datos de forma directa a los archivos de datos en Amazon Simple Storage Service. Para obtener más información, consulte Descarga de datos en Amazon Redshift.

- Establezca el parámetro de búsqueda de JDBC en la aplicación cliente. Si utiliza una conexión JDBC y se encuentra con errores de falta de memoria del lado del cliente, puede habilitar el cliente para recuperar conjuntos de resultados en lotes más pequeños al configurar el parámetro de tamaño de búsqueda de JDBC. Para obtener más información, consulte Configuración del parámetro de tamaño de búsqueda de la JDBC.

## Ejemplos de DECLARE CURSOR

En el siguiente ejemplo, se declara un cursor denominado LOLLAPALOOZA para seleccionar información de ventas para el evento Lollapalooza y, luego, se recuperan filas del conjunto de resultados a través del cursor:

```
-- Begin a transaction

begin;

-- Declare a cursor

declare lollapalooza cursor for
select eventname, starttime, pricepaid/qtysold as costperticket, qtysold
from sales, event
where sales.eventid = event.eventid
and eventname='Lollapalooza';

-- Fetch the first 5 rows in the cursor lollapalooza:

fetch forward 5 from lollapalooza;

  eventname   |      starttime      | costperticket | qtysold
--------------+---------------------+---------------+---------
 Lollapalooza | 2008-05-01 19:00:00 |   92.00000000 |       3
 Lollapalooza | 2008-11-15 15:00:00 |  222.00000000 |       2
 Lollapalooza | 2008-04-17 15:00:00 |  239.00000000 |       3
 Lollapalooza | 2008-04-17 15:00:00 |  239.00000000 |       4
 Lollapalooza | 2008-04-17 15:00:00 |  239.00000000 |       1
(5 rows)

-- Fetch the next row:

fetch next from lollapalooza;

  eventname   |      starttime      | costperticket | qtysold
```

```
-------------+---------------------+--------------+---------
 Lollapalooza | 2008-10-06 14:00:00 |  114.00000000 |       2

-- Close the cursor and end the transaction:

close lollapalooza;
commit;
```

El siguiente ejemplo realiza un bucle en un refcursor con todos los resultados de una tabla:

```
CREATE TABLE tbl_1 (a int, b int);
INSERT INTO tbl_1 values (1, 2),(3, 4);

CREATE OR REPLACE PROCEDURE sp_cursor_loop() AS $$
DECLARE
    target record;
    curs1 cursor for select * from tbl_1;
BEGIN
    OPEN curs1;
    LOOP
        fetch curs1 into target;
        exit when not found;
        RAISE INFO 'a %', target.a;
    END LOOP;
    CLOSE curs1;
END;
$$ LANGUAGE plpgsql;

CALL sp_cursor_loop();

SELECT message
   from svl_stored_proc_messages
   where querytxt like 'CALL sp_cursor_loop()%';

  message
----------
      a 1
      a 3
```

# DELETE

Elimina filas de tablas.

> **Note**
>
> El tamaño máximo de una instrucción SQL es de 16 MB.

## Sintaxis

```
[ WITH [RECURSIVE] common_table_expression [, common_table_expression , ...] ]
DELETE [ FROM ] { table_name | materialized_view_name }
    [ { USING } table_name, ... ]
    [ WHERE condition ]
```

## Parámetros

### Cláusula WITH

Se trata de una cláusula opcional que especifica una o más common-table-expressions. Consulte Cláusula WITH.

### FROM

La palabra clave FROM es opcional, excepto cuando se especifica la cláusula USING. Las instrucciones `delete from event;` y `delete event;` son operaciones equivalentes que eliminan todas las filas de la tabla EVENT.

> **Note**
>
> Para eliminar todas las filas de una tabla, TRUNCATE la tabla. TRUNCATE es mucho más eficiente que DELETE y no requiere VACUUM ni ANALYZE. No obstante, tenga en cuenta que TRUNCATE confirma la transacción en la que se ejecuta.

### table_name

Una tabla temporal o persistente. Solo el propietario de la tabla o un usuario con el privilegio DELETE en la tabla puede eliminar filas de la tabla.

Considere utilizar el comando TRUNCATE para las operaciones de eliminación incompletas rápidas en tablas grandes; para ello, consulte TRUNCATE.

> **Note**
>
> Después de eliminar una gran cantidad de filas de una tabla:
>
> - Limpie la tabla para recuperar espacio de almacenamiento y reordene las filas.
> - Analice la tabla para actualizar las estadísticas para el planificador de consultas.

materialized_view_name

Una vista materializada. La instrucción DELETE funciona en una vista materializada utilizada para Ingesta de streaming a una vista materializada. Solo el propietario de la vista materializada o un usuario con privilegio DELETE sobre la vista materializada puede eliminar filas de ella.

No puede ejecutar DELETE en una vista materializada para la ingesta de streaming con una política de seguridad a nivel de fila (RLS) que no tenga el permiso IGNORE RLS concedido al usuario. Hay una excepción a esto: Si el usuario que realiza la operación DELETE tiene concedida la opción IGNORE RLS, esta se ejecuta correctamente. Para obtener más información, consulte Propiedad y administración de la política de RLS.

USING table_name (nombre_de_tabla), ...

La palabra clave USING se utiliza para presentar una lista de tabla cuando se hace referencia a tablas adicionales en la condición de la cláusula WHERE. Por ejemplo, la siguiente instrucción elimina todas las filas de la tabla EVENT que cumplen con la condición de combinación sobre las tablas EVENT y SALES. La tabla SALES debe tener un nombre designado explícitamente en la lista FROM:

```
delete from event using sales where event.eventid=sales.eventid;
```

Si repite el nombre de la tabla de destino en la cláusula USING, la operación DELETE ejecuta una combinación automática. Puede usar una subconsulta en la cláusula WHERE en lugar de la sintaxis USING como una manera alternativa para escribir la misma consulta.

WHERE condition

Cláusula opcional que limita la eliminación de filas a aquellas que coinciden con la condición. Por ejemplo, la condición puede ser una restricción en una columna, una condición de combinación o una condición basada en el resultado de una consulta. La consulta puede hacer referencia a tablas diferentes a la tabla de origen del comando DELETE. Por ejemplo:

```
delete from t1
where col1 in(select col2 from t2);
```

Si no se especifica una condición, se eliminan todas las filas de la tabla.

## Ejemplos

Elimine todas las filas de la tabla CATEGORY:

```
delete from category;
```

Elimine las filas con valores CATID entre 0 y 9 de la tabla CATEGORY:

```
delete from category
where catid between 0 and 9;
```

Elimine las filas de la tabla LISTING cuyos valores SELLERID no existan en la tabla SALES:

```
delete from listing
where listing.sellerid not in(select sales.sellerid from sales);
```

Las siguientes dos consultas eliminan una fila de la tabla CATEGORY en función de una combinación con la tabla EVENT y una restricción adicional en la columna CATID:

```
delete from category
using event
where event.catid=category.catid and category.catid=9;
```

```
delete from category
where catid in
(select category.catid from category, event
where category.catid=event.catid and category.catid=9);
```

La siguiente consulta elimina todas las filas de la vista materializada `mv_cities`. El nombre de vista materializada en este ejemplo es una muestra:

```
delete from mv_cities;
```

# DESC DATASHARE

Muestra una lista de los objetos de la base de datos dentro de un datashare que se agregan a él con ALTER DATASHARE. Amazon Redshift muestra los nombres, las bases de datos, los esquemas y los tipos de tablas, vistas y funciones.

Puede encontrar información adicional sobre los objetos del recurso compartido de datos utilizando las vistas del sistema. Para obtener más información, consulte SVV_DATASHARE_OBJECTS y SVV_DATASHARES.

## Sintaxis

```
DESC DATASHARE datashare_name [ OF [ ACCOUNT account_id ] NAMESPACE namespace_guid ]
```

## Parámetros

datashare_name

Se trata del nombre del datashare.

NAMESPACE namespace_guid

Se trata de un valor que especifica el espacio de nombres que utiliza el datashare. Cuando ejecute DESC DATASHARE como administrador de un clúster consumidor, especifique el parámetro NAMESPACE para ver los datashares entrantes.

ACCOUNT account_id

Se trata de un valor que especifica la cuenta a la que pertenece el datashare.

## Notas de uso

Como administrador de una cuenta consumidora, cuando ejecute DESC DATASHARE para ver los datashares entrantes en la cuenta de AWS, especifique la opción NAMESPACE. Cuando ejecute DESC DATASHARE para ver los datashares entrantes en las cuentas de AWS, especifique las opciones ACCOUNT y NAMESPACE.

## Ejemplos

En el siguiente ejemplo, se muestra la información de los datashares salientes en un clúster productor.

```
DESC DATASHARE salesshare;

producer_account |          producer_namespace           | share_type  | share_name   |
 object_type |        object_name           |  include_new
-----------------+--------------------------------------+-------------+--------------
+-------------+------------------------------+--------------
 123456789012    | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d  | OUTBOUND    |  salesshare  |
 TABLE       | public.tickit_sales_redshift |
 123456789012    | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d  | OUTBOUND    |  salesshare  |
 SCHEMA      | public                       |   t
```

En el siguiente ejemplo, se muestra la información de los datashares entrantes en un clúster consumidor.

```
DESC DATASHARE salesshare of ACCOUNT '123456789012' NAMESPACE
 '13b8833d-17c6-4f16-8fe4-1a018f5ed00d';

 producer_account |          producer_namespace          | share_type | share_name |
 object_type |         object_name          |  include_new
-----------------+--------------------------------------+------------+------------
+-------------+------------------------------+--------------
 123456789012     | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    | salesshare |
 table       | public.tickit_sales_redshift |
 123456789012     | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d | INBOUND    | salesshare |
 schema      | public                       |
(2 rows)
```

# DESC IDENTITY PROVIDER

Muestra información sobre un proveedor de identidades. Solo un superusuario puede crear un proveedor de identidades.

## Sintaxis

```
DESC IDENTITY PROVIDER identity_provider_name
```

## Parámetros

identity_provider_name

El nombre del proveedor de identidades.

### Ejemplo

En el siguiente ejemplo se muestra información acerca del proveedor de identidades.

```
DESC IDENTITY PROVIDER azure_idp;
```

Resultados de ejemplo.

```
  uid   |   name    | type  |              instanceid              | namespc |

                                          params

                           | enabled
--------+-----------+-------+--------------------------------------+---------
+-----------------------------------------------------------------------------------------------------------------
+---------
 126692 | azure_idp | azure | e40d4bb2-7670-44ae-bfb8-5db013221d73 | aad     |
 {"issuer":"https://login.microsoftonline.com/e40d4bb2-7670-44ae-bfb8-5db013221d73/
v2.0", "client_id":"871c010f-5e61-4fb1-83ac-98610a7e9110", "client_secret":'',
 "audience":["https://analysis.windows.net/powerbi/connector/AmazonRedshift", "https://
analysis.windows.net/powerbi/connector/AWSRDS"]} | t
(1 row)
```

## DETACH MASKING POLICY

Desconecta una política de enmascaramiento dinámico de datos ya adjunta de una columna. Para obtener más información sobre el enmascaramiento dinámico de datos, consulte Enmascaramiento de datos dinámico.

Los superusuarios y los usuarios o roles que tienen el rol sys:secadmin pueden desconectar una política de enmascaramiento.

### Sintaxis

```
DETACH MASKING POLICY policy_name
   ON { table_name }
   ( output_column_names )
   FROM { user_name | ROLE role_name | PUBLIC };
```

### Parámetros

policy_name

El nombre de la política de enmascaramiento que se va a desconectar.

table_name

El nombre de la tabla de la que se puede desconectar la tabla de enmascaramiento.

output_column_names

Nombres de las columnas a las que se asoció la política de enmascaramiento.

user_name

Nombre del usuario al que se asoció la política de enmascaramiento.

Solo puede establecer user_name, role_name o PUBLIC en una sola instrucción DETACH MASKING POLICY.

role_name

Nombre del rol al que se asoció la política de enmascaramiento.

Solo puede establecer user_name, role_name o PUBLIC en una sola instrucción DETACH MASKING POLICY.

PUBLIC

Muestra que la política se asoció a todos los usuarios de la tabla.

Solo puede establecer user_name, role_name o PUBLIC en una sola instrucción DETACH MASKING POLICY.

## DETACH RLS POLICY

Desconecte una política de seguridad de la fila en una tabla de uno o más usuarios o roles.

Los superusuarios y los usuarios o roles que tienen el rol `sys:secadmin` pueden desconectar una política.

### Sintaxis

```
DETACH RLS POLICY policy_name ON [TABLE] table_name [, ...]
```

```
FROM { user_name | ROLE role_name | PUBLIC } [, ...]
```

### Parámetros

policy_name

El nombre de la política de .

ON [TABLE] table_name [, ...]

La tabla o la vista de la que está desconectada la política de seguridad de la fila.

FROM { user_name | ROLE role_name | PUBLIC} [, ...]

Especifica si la política está desconectada de uno o más usuarios o roles especificados.

### Notas de uso

Al trabajar con la instrucción DETACH RLS POLICY, observe lo siguiente:

- Puede desconectar una política de una relación, usuario, rol o público.

### Ejemplos

En el ejemplo siguiente, se desconecta una política en una tabla de un rol.

```
DETACH RLS POLICY policy_concerts ON tickit_category_redshift FROM ROLE analyst, ROLE
 dbadmin;
```

## DROP DATABASE

Elimina una base de datos.

No se puede ejecutar DROP DATABASE en un bloque de transacción (BEGIN … END). Para obtener más información acerca de las transacciones, consulte Aislamiento serializable.

### Sintaxis

```
DROP DATABASE database_name
```

## Parámetros

database_name

Nombre de la base de datos que se eliminará. No se pueden eliminar las bases de datos dev, padb_harvest, template0, template1 o sys:internal, ni tampoco se puede eliminar la base de datos actual.

Para eliminar una base de datos externa, elimine el esquema externo. Para obtener más información, consulte DROP SCHEMA.

## Notas de uso de DROP DATABASE

Cuando utilice la instrucción DROP DATABASE, tenga en cuenta lo siguiente:

- En general, se recomienda que no elimine una base de datos que contenga un recurso compartido de datos de AWS Data Exchange mediante la instrucción DROP DATABASE. Si lo hace, las Cuentas de AWS que tengan acceso al recurso compartido de datos perderán el acceso. Realizar este tipo de alteración puede infringir los términos del producto de datos en AWS Data Exchange.

  El siguiente ejemplo muestra un error cuando se elimina una base de datos que contiene un recurso compartido de datos de AWS Data Exchange.

```
DROP DATABASE test_db;
ERROR:   Drop of database test_db that contains ADX-managed datashare(s)
 requires session variable datashare_break_glass_session_var to be set to value
 'ce8d280c10ad41'
```

  Para permitir la eliminación de la base de datos, configure la siguiente variable y ejecute de nuevo la instrucción DROP DATABASE.

```
SET datashare_break_glass_session_var to 'ce8d280c10ad41';
```

```
DROP DATABASE test_db;
```

  En este caso, Amazon Redshift genera un valor aleatorio único para establecer la variable de sesión para permitir DROP DATABASE para una base de datos que contiene un recurso compartido de datos de AWS Data Exchange.

### Ejemplos

En el siguiente ejemplo, se elimina una base de datos denominada TICKIT_TEST:

```
drop database tickit_test;
```

## DROP DATASHARE

Elimina un datashare. Este comando no es reversible.

Solo un superusuario o el propietario de datashare puede eliminar uno de estos.

### Privilegios necesarios

Los siguientes privilegios son necesarios para DROP DATASHARE:

- Superusuario
- Usuarios con el privilegio DROP DATASHARE
- Propietario del recurso compartido de datos

### Sintaxis

```
DROP DATASHARE datashare_name;
```

### Parámetros

datashare_name

Se trata del nombre del datashare que se eliminará.

### Notas de uso de DROP DATASHARE

Cuando utilice la instrucción DROP DATASHARE, tenga en cuenta lo siguiente:

- En general, le recomendamos que no elimine un recurso compartido de datos de AWS Data Exchange mediante la instrucción DROP DATASHARE. Si lo hace, las Cuentas de AWS que tengan acceso al recurso compartido de datos perderán el acceso. Realizar este tipo de alteración puede infringir los términos del producto de datos en AWS Data Exchange.

El siguiente ejemplo muestra un error cuando se elimina un recurso compartido de datos de AWS Data Exchange.

```
DROP DATASHARE salesshare;
ERROR:  Drop of ADX-managed datashare salesshare requires session variable
 datashare_break_glass_session_var to be set to value '620c871f890c49'
```

Para permitir la eliminación del recurso compartido de datos de AWS Data Exchange, configure la siguiente variable y ejecute de nuevo la instrucción DROP DATASHARE.

```
SET datashare_break_glass_session_var to '620c871f890c49';
```

```
DROP DATASHARE salesshare;
```

En este caso, Amazon Redshift genera un valor aleatorio único para establecer la variable de sesión para permitir DROP DATASHARE para un recurso compartido de datos de AWS Data Exchange.

### Ejemplos

En el siguiente ejemplo, se elimina un recurso compartido de datos denominado `salesshare`.

```
DROP DATASHARE salesshare;
```

## DROP EXTERNAL VIEW

Elimina una vista externa de la base de datos. Al eliminar una vista externa, se elimina de todos los motores de SQL a los que está asociada la vista, como Amazon Athena y Amazon EMR Spark. Este comando no se puede revertir. Para obtener más información sobre las vistas del Catálogo de datos, consulte Vistas de AWS Glue Data Catalog.

### Sintaxis

```
DROP EXTERNAL VIEW schema_name.view_name [ IF EXISTS ]
{catalog_name.schema_name.view_name | awsdatacatalog.dbname.view_name |
 external_schema_name.view_name}
```

### Parámetros

schema_name.view_name

Es el esquema asociado a la base de datos de AWS Glue, seguido del nombre de la vista.

IF EXISTS

Elimina la vista solo si existe.

catalog_name.schema_name.view_name | awsdatacatalog.dbname.view_name | external_schema_name.view_name

Es la notación del esquema que se utilizará al eliminar la vista. Puede especificar el uso de AWS Glue Data Catalog, una base de datos de Glue que haya creado o un esquema externo que haya creado. Consulte CREATE DATABASE y CREATE EXTERNAL SCHEMA para obtener más información.

query_definition

Es la definición de la consulta SQL que Amazon Redshift ejecuta para modificar la vista.

### Ejemplos

El siguiente ejemplo elimina una vista del catálogo de datos denominada sample_schema.glue_data_catalog_view.

```
DROP EXTERNAL VIEW sample_schema.glue_data_catalog_view IF EXISTS
```

## DROP FUNCTION

Elimina una función definida por el usuario (UDF) de la base de datos. Se debe especificar la firma de la función, o una lista de tipos de datos de argumento, ya que puede haber varias funciones con el mismo nombre pero con diferentes firmas. No se puede eliminar una función integrada de Amazon Redshift.

Este comando no es reversible.

### Privilegios necesarios

Los siguientes privilegios son necesarios para DROP FUNCTION:

- Superusuario
- Usuarios con el privilegio DROP FUNCTION
- Propietario de la función

## Sintaxis

```
DROP FUNCTION name
( [arg_name] arg_type   [, ...] )
[ CASCADE | RESTRICT ]
```

## Parámetros

name

El nombre de la función que se eliminará.

arg_name (nombre_de_argumento)

El nombre de un argumento de entrada. DROP FUNCTION ignora nombres de argumento ya que solo los tipos de datos de argumento son necesarios para determinar la identidad de la función.

arg_type

El tipo de datos del argumento de entrada. Puede proporcionar una lista de valores separados por coma con un máximo de 32 tipos de datos.

CASCADE

Palabra clave que especifica que se deben eliminar automáticamente los objetos que dependen de una función, como las vistas.

Para crear una vista que no dependa de una función, incluya la cláusula WITH NO SCHEMA BINDING en la definición de vista. Para obtener más información, consulte CREATE VIEW.

RESTRICT

Se trata de la palabra clave que especifica que, si hay objetos que dependen de la función, no se debe eliminar la función y devolver un mensaje. Esta acción es la predeterminada.

## Ejemplos

En el siguiente ejemplo, se elimina la función denominada `f_sqrt`:

```
drop function f_sqrt(int);
```

Para eliminar una función que tiene dependencias, use la opción CASCADE, como se muestra en el siguiente ejemplo:

```
drop function f_sqrt(int)cascade;
```

# DROP GROUP

Elimina un grupo de usuarios. Este comando no es reversible. Este comando no elimina a los usuarios individuales de un grupo.

Consulte DROP USER para eliminar un usuario individual.

## Sintaxis

```
DROP GROUP name
```

## Parámetro

name

Nombre del grupo de usuarios que se eliminará.

## Ejemplo

En el siguiente ejemplo, se elimina el grupo de usuarios `guests`:

```
DROP GROUP guests;
```

No se puede eliminar un grupo si el grupo tiene privilegios en un objeto. Si intenta eliminar dicho grupo, recibirá el siguiente error.

```
ERROR: group "guests" can't be dropped because the group has a privilege on some object
```

Si el grupo tiene privilegios para un objeto, debe eliminar los privilegios antes de eliminar el grupo. Para buscar los objetos para los que el grupo `guests` tiene privilegios, utilice el siguiente ejemplo. Para obtener más información sobre la vista de metadatos utilizada en el ejemplo, consulte SVV_RELATION_PRIVILEGES.

```
SELECT DISTINCT namespace_name, relation_name, identity_name, identity_type
FROM svv_relation_privileges
WHERE identity_type='group' AND identity_name='guests';

+----------------+---------------+---------------+---------------+
| namespace_name | relation_name | identity_name | identity_type |
+----------------+---------------+---------------+---------------+
| public         | table1        | guests        | group         |
+----------------+---------------+---------------+---------------+
| public         | table2        | guests        | group         |
+----------------+---------------+---------------+---------------+
```

En el siguiente ejemplo, se eliminan todos los privilegios en todas las tablas del esquema `public` del grupo de usuarios `guests` y, luego, se elimina el grupo.

```
REVOKE ALL ON ALL TABLES IN SCHEMA public FROM GROUP guests;
DROP GROUP guests;
```

# DROP IDENTITY PROVIDER

Elimina un proveedor de identidades. Este comando no es reversible. Solo un superusuario puede eliminar un proveedor de identidades.

## Sintaxis

```
DROP IDENTITY PROVIDER identity_provider_name [ CASCADE ]
```

## Parámetros

identity_provider_name

Nombre del proveedor de identidades que se va a eliminar.

CASCADE

Elimina los usuarios y roles asociados al proveedor de identidades cuando se elimina.

## Ejemplo

En el siguiente ejemplo se elimina el proveedor de identidades oauth_provider.

```
DROP IDENTITY PROVIDER oauth_provider;
```

Si elimina el proveedor de identidades, es posible que algunos usuarios no puedan iniciar sesión o utilizar herramientas de cliente configuradas para utilizar el proveedor de identidades.

# DROP LIBRARY

Elimina una biblioteca de Python personalizada de la base de datos. Solo el propietario de la biblioteca o un superusuario puede eliminar una biblioteca.

DROP LIBRARY no se puede ejecutar dentro de un bloque de transacción (BEGIN … END). Para obtener más información acerca de las transacciones, consulte Aislamiento serializable.

Este comando no es reversible. El comando DROP LIBRARY se confirma inmediatamente. Si una UDF que depende de la biblioteca se ejecuta simultáneamente, es posible que la UDF falle, incluso si la UDF se ejecuta dentro de una transacción.

Para obtener más información, consulte CREATE LIBRARY.

## Privilegios necesarios

Los siguientes privilegios son necesarios para DROPLIBRARY:

- Superusuario
- Usuarios con el privilegio DROP LIBRARY
- Propietario de la biblioteca

## Sintaxis

```
DROP LIBRARY library_name
```

## Parámetros

library_name

El nombre de la biblioteca.

# DROP MASKING POLICY

Elimina una política de enmascaramiento dinámico de datos de todas las bases de datos. No puede eliminar una política de enmascaramiento si sigue adjunta a una o más tablas. Para obtener más información sobre el enmascaramiento dinámico de datos, consulte Enmascaramiento de datos dinámico.

Los superusuarios y los usuarios o roles que tienen el rol sys:secadmin pueden eliminar una política de enmascaramiento.

## Sintaxis

```
DROP MASKING POLICY policy_name;
```

## Parámetros

policy_name

El nombre de la política de enmascaramiento que se va a eliminar.

# DROP MODEL

Quita un modelo de la base de datos. Solo el propietario del modelo o un superusuario puede quitarlo.

DROP MODEL también elimina toda función de predicción asociada que se obtenga de este modelo, todos los artefactos de Amazon Redshift relacionados con el modelo y todos los datos de Amazon S3 relacionados con el modelo. Mientras el modelo se está entrenando en Amazon SageMaker AI, DROP MODEL cancelará esas operaciones.

Este comando no es reversible. El comando DROP MODEL se confirma de forma inmediata.

## Permisos necesarios

Los siguientes permisos son necesarios para DROP MODEL:

- Superusuario
- Usuarios con el permiso DROP MODEL
- Propietario del modelo

- Propietario del esquema

### Sintaxis

```
DROP MODEL [ IF EXISTS ] model_name
```

### Parámetros

IF EXISTS

Se trata de una cláusula que indica que, si el esquema especificado ya existe, el comando no debe realizar cambios y debe devolver un mensaje en el que se indique que el esquema existe.

model_name

El nombre del modelo. El nombre del modelo en un esquema debe ser único.

### Ejemplos

En el siguiente ejemplo, se elimina el modelo demo_ml.customer_churn.

```
DROP MODEL demo_ml.customer_churn
```

## DROP MATERIALIZED VIEW

Retira una vista materializada.

Para obtener más información acerca de las vistas materializadas, consulte Vistas materializadas en Amazon Redshift.

### Sintaxis

```
DROP MATERIALIZED VIEW [ IF EXISTS ] mv_name [ CASCADE | RESTRICT ]
```

### Parámetros

IF EXISTS

Cláusula que especifica que hay que comprobar si existe la vista materializada nombrada. Si la vista materializada no existe, el comando `DROP MATERIALIZED VIEW` le mostrará un mensaje

de error. Esta cláusula es útil cuando se realiza scripting, para que el script no produzca un error si se elimina una vista materializada inexistente.

mv_name

El nombre de la vista materializada que desea eliminar.

CASCADE

Cláusula que indica que se eliminen automáticamente los objetos de los que depende la vista materializada, como otras vistas.

RESTRICT

Cláusula que indica que no se elimine la vista materializada si algún objeto depende de ella. Esta es la opción predeterminada.

### Notas de uso

Solo el propietario de una vista materializada puede utilizar `DROP MATERIALIZED VIEW` en esa vista. Las excepciones pueden ser un superusuario o un usuario al que se le hayan otorgado específicamente privilegios DROP.

Al escribir una instrucción drop para una vista materializada y una vista con un nombre coincidente, se produce un error que indica que debe utilizar DROP VIEW. Se produce un error incluso en el caso de que utilice `DROP MATERIALIZED VIEW IF EXISTS`.

### Ejemplo

El siguiente ejemplo elimina la vista materializada de `tickets_mv`.

```
DROP MATERIALIZED VIEW tickets_mv;
```

## DROP PROCEDURE

Elimina un procedimiento. Para eliminar un procedimiento se necesitan el nombre del procedimiento y los tipos de datos del argumento de entrada (firma). De manera opción, puede incluir los tipos de datos del argumento completo, incluidos los argumentos OUT. Para buscar la firma de un procedimiento, utilice el comando SHOW PROCEDURE. Para obtener más información acerca de las firmas de procedimientos, consulte PG_PROC_INFO.

## Privilegios necesarios

Los siguientes privilegios son necesarios para DROP PROCEDURE:

- Superusuario
- Usuarios con el privilegio DROP PROCEDURE
- Propietario del procedimiento

## Sintaxis

```
DROP PROCEDURE sp_name ( [ [ argname ] [ argmode ] argtype [, ...] ] )
```

## Parámetros

sp_name

El nombre del procedimiento que debe eliminarse.

argname

El nombre de un argumento de entrada. DROP PROCEDURE ignora nombres de argumento ya que solo los tipos de datos de argumento son necesarios para determinar la identidad del procedimiento.

argmode

El modo de un argumento, que puede ser IN, OUT o INOUT. Los argumentos OUT son opcionales porque no se utilizan para identificar un procedimiento almacenado.

argtype

El tipo de datos del argumento de entrada. Para ver una lista de los tipos de datos admitidos, consulte Tipos de datos.

## Ejemplos

El siguiente ejemplo elimina un procedimiento almacenado denominado `quarterly_revenue`.

```
DROP PROCEDURE quarterly_revenue(volume INOUT bigint, at_price IN numeric,result OUT
 int);
```

# DROP RLS POLICY

Elimina una política de seguridad de la fila de todas las tablas en todas las bases de datos.

Los superusuarios y los usuarios o roles que tienen el rol sys:secadmin pueden eliminar una política.

## Sintaxis

```
DROP RLS POLICY [ IF EXISTS ] policy_name [ CASCADE | RESTRICT ]
```

## Parámetros

IF EXISTS

Una cláusula que indica si la política especificada ya existe.

policy_name

El nombre de la política de .

CASCADE

Una cláusula que indica que se debe desconectar automáticamente la política de todas las tablas adjuntas antes de eliminarla.

RESTRICT

Una cláusula que indica que no se debe eliminar la política cuando está adjuntada a algunas tablas. Esta es la opción predeterminada.

## Ejemplos

En el siguiente ejemplo, se elimina la política de seguridad de la fila.

```
DROP RLS POLICY policy_concerts;
```

# DROP ROLE

Elimina un rol de una base de datos. Solo pueden eliminar un rol el propietario del rol que lo haya creado, un usuario con la opción WITH ADMIN o un superusuario.

No se puede eliminar un rol que esté concedido a un usuario o si otro rol depende de él.

## Privilegios necesarios

Los siguientes privilegios son necesarios para DROP ROLE:

- Superusuario
- Propietario del rol que sea el usuario que haya creado el rol o un usuario al que se le haya concedido el rol con el privilegio WITH ADMIN OPTION.

## Sintaxis

```
DROP ROLE role_name [ FORCE | RESTRICT ]
```

## Parámetros

role_name

Nombre del rol

[ FORCE | RESTRICT ]

La configuración predeterminada es RESTRICT. Amazon Redshift genera un error cuando se intenta eliminar un rol que ha heredado otro rol. Utilice FORCE para eliminar todas las asignaciones de roles, si existe alguna.

## Ejemplos

El siguiente ejemplo elimina el rol `sample_role`.

```
DROP ROLE sample_role FORCE;
```

El siguiente ejemplo intenta eliminar el rol sample_role1 que se ha concedido a un usuario con la opción RESTRICT predeterminada.

```
CREATE ROLE sample_role1;
GRANT ROLE sample_role1 TO user1;
DROP ROLE sample_role1;
ERROR:  cannot drop this role since it has been granted on a user
```

Para eliminar correctamente el rol sample_role1 que se haya concedido a un usuario, utilice la opción FORCE.

```
DROP ROLE sample_role1 FORCE;
```

El siguiente ejemplo intenta eliminar el rol sample_role2 que tiene otro rol dependiente de él con la opción RESTRICT predeterminada.

```
CREATE ROLE sample_role1;
CREATE ROLE sample_role2;
GRANT ROLE sample_role1 TO sample_role2;
DROP ROLE sample_role2;
ERROR:  cannot drop this role since it depends on another role
```

Para eliminar correctamente el rol sample_role2 que tenga otro rol que dependa de él, utilice la opción FORCE.

```
DROP ROLE sample_role2 FORCE;
```

# DROP SCHEMA

Elimina un esquema. Para un esquema externo, también puede eliminar la base de datos externa asociada con el esquema. Este comando no es reversible.

## Privilegios necesarios

Los siguientes privilegios son necesarios para DROP SCHEMA:

- Superusuario
- Propietario del esquema
- Usuarios con el privilegio DROP SCHEMA

## Sintaxis

```
DROP SCHEMA [ IF EXISTS ] name [, ...]
[ DROP EXTERNAL DATABASE ]
[ CASCADE | RESTRICT ]
```

## Parámetros

IF EXISTS

Cláusula que indica que si el esquema especificado no existe, el comando no debe realizar cambios y debe devolver un mensaje en el que se indique que el esquema no existe, en lugar de terminar con un error.

Esta cláusula es útil cuando se realiza scripting, para que el script no produzca un error si DROP SCHEMA se ejecuta contra un esquema que no existe.

name

Nombres de los esquemas que eliminar. Puede especificar múltiples nombres de esquema separados por comas.

DROP EXTERNAL DATABASE

Cláusula que indica que si se elimina un esquema externo, se elimina la base de datos externa asociada con el esquema externo, si existe una. Si no existe ninguna base de datos externa, el comando devuelve un mensaje que indica que no existe ninguna base de datos externa. Si se eliminan múltiples esquemas externos, se eliminan todas las bases de datos externas asociadas con los esquemas especificados.

Si una base de datos externa contiene objetos dependientes como son las tablas, incluya la opción CASCADE para eliminar también los objetos dependientes.

Cuando se elimina una base de datos externa, la base de datos también se elimina de cualquier otro esquema externo asociado con la base de datos. Las tablas definidas en otros esquemas externos usando la base de datos también se eliminan.

DROP EXTERNAL DATABASE no es compatible con bases de datos externas almacenadas en un almacén de metadatos de HIVE.

CASCADE

Palabra clave que indica que se deben eliminar automáticamente todos los objetos del esquema. Si se especifica DROP EXTERNAL DATABASE, también se eliminan todos los objetos de la base de datos externa.

RESTRICT

Palabra clave que indica que no se debe eliminar un esquema o una base de datos externa si contiene objetos. Esta acción es la predeterminada.

### Ejemplo

En el siguiente ejemplo, se elimina un esquema denominado S_SALES. Este ejemplo usa RESTRICT como mecanismo de seguridad para que no se elimine el esquema si contiene objetos. En este caso, necesita eliminar los objetos del esquema antes de eliminar el esquema.

```
drop schema s_sales restrict;
```

En el siguiente ejemplo, se eliminan un esquema denominado S_SALES y todos los objetos que dependen de ese esquema.

```
drop schema s_sales cascade;
```

En el siguiente ejemplo, se elimina el esquema S_SALES si existe. Si no existe, no se realiza ninguna acción y se devuelve un mensaje.

```
drop schema if exists s_sales;
```

En el siguiente ejemplo, se elimina un esquema externo denominado S_SPECTRUM y la base de datos externa asociada con él. Este ejemplo usa RESTRICT para que no se elimine el esquema ni la base de datos si contienen objetos. En este caso, necesita eliminar los objetos dependientes antes de eliminar el esquema y la base de datos.

```
drop schema s_spectrum drop external database restrict;
```

En el siguiente ejemplo, se eliminan múltiples esquemas y las bases de datos asociadas con ellos, junto con todos los objetos dependientes.

```
drop schema s_sales, s_profit, s_revenue drop external database cascade;
```

## DROP TABLE

Elimina una tabla de una base de datos.

Si desea borrar las filas de una tabla sin eliminar la tabla, use el comando DELETE o TRUNCATE.

DROP TABLE elimina las restricciones que existen en la tabla de destino. Puede eliminar varias tablas con un solo comando DROP TABLE.

No se puede ejecutar DROP TABLE con una tabla externa en una transacción (BEGIN … END). Para obtener más información acerca de las transacciones, consulte Aislamiento serializable.

Para encontrar un ejemplo en el que se conceda el privilegio DROP a un grupo, consulte Ejemplos de GRANT.

## Privilegios necesarios

Los siguientes privilegios son necesarios para DROP TABLE:

- Superusuario
- Usuarios con el privilegio DROP TABLE
- Propietario de la tabla con el privilegio USAGE en el esquema

## Sintaxis

```
DROP TABLE [ IF EXISTS ] name [, ...] [ CASCADE | RESTRICT ]
```

## Parámetros

IF EXISTS

Cláusula que indica que si la tabla especificada no existe, el comando no debe realizar cambios y debe devolver un mensaje en el que se indique que la tabla no existe, en lugar de terminar con un error.

Esta cláusula es útil cuando se realiza scripting, para que el script no produzca un error si DROP TABLE se ejecuta contra una tabla que no existe.

name

Nombre de la tabla que se eliminará.

CASCADE

Cláusula que indica que se deben eliminar automáticamente los objetos que dependen de la tabla, como las vistas.

Para crear una vista que no dependa de otros objetos de base de datos, como vistas y tablas, incluya la cláusula WITH NO SCHEMA BINDING en la definición de la vista. Para obtener más información, consulte CREATE VIEW.

RESTRICT

Cláusula que indica que no se debe eliminar la tabla si tiene objetos dependientes. Esta acción es la predeterminada.

## Ejemplos

Eliminar una tabla sin dependencias

En el siguiente ejemplo, se crea y se elimina una tabla denominada FEEDBACK que no tiene dependencias:

```
create table feedback(a int);

drop table feedback;
```

Si una tabla contiene columnas que se referencian en las vistas o en otras tablas, Amazon Redshift muestra un mensaje similar al siguiente.

```
Invalid operation: cannot drop table feedback because other objects depend on it
```

Eliminar dos tablas simultáneamente

El siguiente conjunto de comandos crea una tabla FEEDBACK y una tabla BUYERS y, luego, elimina ambas tablas con un único comando:

```
create table feedback(a int);

create table buyers(a int);

drop table feedback, buyers;
```

Eliminar una tabla con una dependencia

Los siguientes pasos muestran cómo eliminar una tabla denominada FEEDBACK con el comando CASCADE.

Primero, cree una tabla simple denominada FEEDBACK con el comando CREATE TABLE:

```
create table feedback(a int);
```

Luego, use el comando CREATE VIEW para crear una vista denominada FEEDBACK_VIEW que dependa de la tabla FEEDBACK:

```
create view feedback_view as select * from feedback;
```

En el siguiente ejemplo, se elimina la tabla FEEDBACK y, también, la vista FEEDBACK_VIEW, ya que FEEDBACK_VIEW depende de la tabla FEEDBACK:

```
drop table feedback cascade;
```

Ver las dependencias de una tabla

Para devolver las dependencias de la tabla, utilice el siguiente ejemplo. Sustituya *my_schema* y *my_table* por su propio esquema y tabla.

```
SELECT dependent_ns.nspname as dependent_schema
, dependent_view.relname as dependent_view
, source_ns.nspname as source_schema
, source_table.relname as source_table
, pg_attribute.attname as column_name
FROM pg_depend
JOIN pg_rewrite ON pg_depend.objid = pg_rewrite.oid
JOIN pg_class as dependent_view ON pg_rewrite.ev_class = dependent_view.oid
JOIN pg_class as source_table ON pg_depend.refobjid = source_table.oid
JOIN pg_attribute ON pg_depend.refobjid = pg_attribute.attrelid
    AND pg_depend.refobjsubid = pg_attribute.attnum
JOIN pg_namespace dependent_ns ON dependent_ns.oid = dependent_view.relnamespace
JOIN pg_namespace source_ns ON source_ns.oid = source_table.relnamespace
WHERE
source_ns.nspname = 'my_schema'
AND source_table.relname = 'my_table'
AND pg_attribute.attnum > 0
ORDER BY 1,2
LIMIT 10;
```

Para eliminar *my_table* y sus dependencias, use el siguiente ejemplo. En este ejemplo también se devuelven todas las dependencias de la tabla que se ha eliminado.

```
DROP TABLE my_table CASCADE;
```

```
SELECT dependent_ns.nspname as dependent_schema
, dependent_view.relname as dependent_view
, source_ns.nspname as source_schema
, source_table.relname as source_table
, pg_attribute.attname as column_name
FROM pg_depend
JOIN pg_rewrite ON pg_depend.objid = pg_rewrite.oid
JOIN pg_class as dependent_view ON pg_rewrite.ev_class = dependent_view.oid
JOIN pg_class as source_table ON pg_depend.refobjid = source_table.oid
JOIN pg_attribute ON pg_depend.refobjid = pg_attribute.attrelid
    AND pg_depend.refobjsubid = pg_attribute.attnum
JOIN pg_namespace dependent_ns ON dependent_ns.oid = dependent_view.relnamespace
JOIN pg_namespace source_ns ON source_ns.oid = source_table.relnamespace
WHERE
source_ns.nspname = 'my_schema'
AND source_table.relname = 'my_table'
AND pg_attribute.attnum > 0
ORDER BY 1,2
LIMIT 10;

+------------------+----------------+---------------+--------------+-------------+
| dependent_schema | dependent_view | source_schema | source_table | column_name |
+------------------+----------------+---------------+--------------+-------------+
```

Eliminar una tabla con IF EXISTS

En el siguiente ejemplo, se elimina la tabla FEEDBACK, si existe. Si no existe, no se realiza ninguna acción y se devuelve un mensaje:

```
drop table if exists feedback;
```

# DROP USER

Elimina un usuario de la base de datos. Puede eliminar varios usuarios con un solo comando DROP USER. Debe ser un superusuario de la base de datos o tener el permiso DROP USER para ejecutar este comando.

## Sintaxis

```
DROP USER [ IF EXISTS ] name [, ... ]
```

## Parámetros

IF EXISTS

Cláusula que indica que si el usuario especificado no existe, el comando no debe realizar cambios y debe devolver un mensaje en el que se indique que el usuario no existe, en lugar de terminar con un error.

Esta cláusula es útil cuando se realiza scripting, para que el script no produzca un error si DROP USER se ejecuta en un usuario que no existe.

name

Nombre del usuario que se eliminará. Puede especificar varios usuarios con una coma separando cada nombre de usuario del siguiente.

## Notas de uso

No puede eliminar el usuario denominado `rdsdb` o el usuario administrador de la base de datos, que normalmente se denomina `awsuser` o `admin`.

No puede eliminar un usuario si el usuario es propietario de cualquier objeto de la base de datos, como un esquema, base de datos, tabla o vista, o si el usuario tiene privilegios en una base de datos, tabla, columna o grupo. Si intenta eliminar dicho usuario, recibirá uno de los siguientes errores.

```
ERROR: user "username" can't be dropped because the user owns some object [SQL
 State=55006]

ERROR: user "username" can't be dropped because the user has a privilege on some object
 [SQL State=55006]
```

Para obtener instrucciones detalladas sobre cómo encontrar los objetos que pertenecen a un usuario de base de datos, consulte ¿Cómo soluciono el error "no se puede eliminar al usuario" en Amazon Redshift? en el Centro de conocimiento.

> **Note**
>
> Amazon Redshift verifica solo la base de datos actual antes de eliminar a un usuario. DROP USER no devuelve un error si el usuario es propietario de objetos de la base de datos o si

tiene privilegios en objetos de otra base de datos. Si elimina un usuario que es propietario de objetos en otra base de datos, el propietario de esos objetos se cambia a "desconocido".

Si un usuario es propietario de un objeto, primero elimine el objeto o elija otro usuario como propietario antes de eliminar el usuario original. Si el usuario tiene privilegios para un objeto, elimine los privilegios antes de eliminar el usuario. En el siguiente ejemplo, se muestra cómo se elimina un objeto, se cambia la propiedad de un objeto y se eliminan privilegios antes de eliminar el usuario.

```
drop database dwdatabase;
alter schema dw owner to dwadmin;
revoke all on table dwtable from dwuser;
drop user dwuser;
```

### Ejemplos

En el siguiente ejemplo, se elimina un usuario denominado paulo:

```
drop user paulo;
```

En el siguiente ejemplo, se eliminan dos usuarios, paulo y martha:

```
drop user paulo, martha;
```

En el siguiente ejemplo, se elimina el usuario paulo, si existe. Si no es así, no se realiza ninguna acción y se devuelve un mensaje:

```
drop user if exists paulo;
```

## DROP VIEW

Elimina una vista de la base de datos. Puede eliminar varias vistas con un solo comando DROP VIEW. Este comando no es reversible.

### Privilegios necesarios

Los siguientes privilegios son necesarios para DROP VIEW:

- Superusuario

- Usuarios con el privilegio DROP VIEW
- Propietario de la vista

## Sintaxis

```
DROP VIEW [ IF EXISTS ] name [, ... ] [ CASCADE | RESTRICT ]
```

## Parámetros

IF EXISTS

Cláusula que indica que si la vista especificada no existe, el comando no debe realizar cambios y debe devolver un mensaje en el que se indique que la vista no existe, en lugar de terminar con un error.

Esta cláusula es útil cuando se realiza scripting, para que el script no produzca un error si DROP VIEW se ejecuta contra una vista que no existe.

name

Nombre de la vista que se eliminará.

CASCADE

Cláusula que indica que se deben eliminar automáticamente los objetos que dependen de la vista, como otras vistas.

Para crear una vista que no dependa de otros objetos de base de datos, como vistas y tablas, incluya la cláusula WITH NO SCHEMA BINDING en la definición de la vista. Para obtener más información, consulte CREATE VIEW.

Tenga en cuenta que si incluye CASCADE y el recuento de objetos de base de datos eliminados asciende a diez o más, es posible que su cliente de base de datos no muestre todos los objetos eliminados en los resultados del resumen. Por lo general, esto se debe a que las herramientas de cliente de SQL tienen limitaciones predeterminadas en los resultados devueltos.

RESTRICT

Cláusula que indica que no se debe eliminar la vista si tiene objetos dependientes. Esta acción es la predeterminada.

## Ejemplos

En el siguiente ejemplo, se elimina la vista event.

```
drop view event;
```

Para eliminar una vista que tiene dependencias, use la opción CASCADE. Por ejemplo, supongamos que comenzamos con una tabla denominada EVENT. Luego, creamos la vista eventview de la tabla EVENT a través del comando CREATE VIEW, como se muestra en el siguiente ejemplo:

```
create view eventview as
select dateid, eventname, catid
from event where catid = 1;
```

Ahora, vamos a crear una segunda vista llamada myeventview, que se basa en la primera vista eventview.

```
create view myeventview as
select eventname, catid
from eventview where eventname <> ' ';
```

Se han creado dos vistas hasta el momento: eventview y myeventview.

La vista myeventview es una vista secundaria de la vista principal eventview.

Para eliminar la vista eventview, parece obvio que el comando que debe utilizarse es el siguiente:

```
drop view eventview;
```

Observe, no obstante, que si ejecuta este comando en este caso recibe el siguiente error:

```
drop view eventview;
ERROR: can't drop view eventview because other objects depend on it
HINT: Use DROP ... CASCADE to drop the dependent objects too.
```

Para solucionarlo, ejecute el siguiente comando (como se sugiere en el mensaje de error):

```
drop view eventview cascade;
```

Ambas vistas, eventview y myeventview, ahora se han eliminado correctamente.

En el siguiente ejemplo, se elimina la vista eventview, si existe. Si no es así, no se realiza ninguna acción y se devuelve un mensaje:

```
drop view if exists eventview;
```

# END

Confirma la transacción actual. Realiza exactamente la misma función que el comando COMMIT.

Consulte COMMIT para obtener información detallada.

## Sintaxis

```
END [ WORK | TRANSACTION ]
```

## Parámetros

WORK

Palabra clave opcional.

TRANSACTION

Palabra clave opcional; WORK y TRANSACTION son sinónimos.

## Ejemplos

Los siguientes ejemplos finalizan el bloque de transacción y confirman la transacción:

```
end;
```

```
end work;
```

```
end transaction;
```

Después de cualquiera de estos comandos, Amazon Redshift finaliza el bloque de transacción y confirma los cambios.

# EXECUTE

Ejecuta una instrucción preparada con anterioridad.

## Sintaxis

```
EXECUTE plan_name [ (parameter [, ...]) ]
```

## Parámetros

plan_name

Nombre de la instrucción preparada para ejecutarse.

parameter

El valor real de un parámetro para la instrucción preparada. Debe ser una expresión que aporta un valor de un tipo compatible con el tipo de datos especificado para este parámetro en el comando PREPARE que creó la instrucción preparada.

## Notas de uso

EXECUTE se utiliza para ejecutar una instrucción preparada con anterioridad. Dado que las instrucciones preparadas solo existen durante una sesión, la instrucción preparada debe crearse con una instrucción PREPARE que se ejecuta anteriormente en la sesión actual.

Si la instrucción PREPARE anterior especificó algunos parámetros, se debe transferir un conjunto compatible de parámetros a la instrucción EXECUTE, o bien, Amazon Redshift devuelve un error. A diferencia de las funciones, las instrucciones preparadas no se sobrecargan en función del tipo o de la cantidad de parámetros especificados; el nombre de una instrucción preparada debe ser único en una sesión de base de datos.

Cuando se emite un comando EXECUTE para la instrucción preparada, Amazon Redshift puede optar por revisar el plan de ejecución de la consulta (para mejorar el rendimiento en función de los valores del parámetro especificado) antes de ejecutar la instrucción preparada. Además, para cada ejecución nueva de una instrucción preparada, Amazon Redshift puede revisar el plan de ejecución de la consulta nuevamente en función de los diferentes valores del parámetro especificados con la instrucción EXECUTE. Para examinar el plan de ejecución de la consulta que Amazon Redshift ha elegido para cualquier instrucción EXECUTE, utilice el comando EXPLAIN.

Para obtener ejemplos y más información acerca de la creación y del uso de instrucciones preparadas, consulte PREPARE.

### Véase también

DEALLOCATE, PREPARE

## EXPLAIN

Muestra el plan de ejecución de una instrucción de consulta sin ejecutar la consulta. Para obtener información sobre el flujo de trabajo del análisis de consultas, consulte Flujo de trabajo de análisis de consultas.

### Sintaxis

```
EXPLAIN [ VERBOSE ] query
```

### Parámetros

VERBOSE

Muestra el plan de consulta completo en lugar de solo un resumen.

consulta

Instrucción de consulta que se explicará. La consulta puede ser una instrucción SELECT, INSERT, CREATE TABLE AS, UPDATE o DELETE.

### Notas de uso

El rendimiento de EXPLAIN a menudo se ve afectado por el tiempo que lleva crear tablas temporales. Por ejemplo, una consulta que usa la optimización de subexpresión común requiere que se creen y analicen tablas temporales para devolver el resultado de EXPLAIN. El plan de consulta depende del esquema y las estadísticas de las tablas temporales. Por lo tanto, el comando EXPLAIN para este tipo de consulta puede tardar más tiempo en ejecutarse que lo esperado.

Puede usar EXPLAIN solo para los siguientes comandos:

- SELECT
- SELECT INTO

- CREATE TABLE AS
- INSERT
- UPDATE
- DELETE

El comando EXPLAIN producirá un error si lo utiliza para otros comandos SQL, como operaciones de base de datos o lenguaje de definición de datos (DDL).

Amazon Redshift utiliza los costos unitarios relativos de los resultados de EXPLAIN para elegir un plan de consultas. Amazon Redshift compara los tamaños de varias estimaciones de recursos para determinar el plan.

## Pasos de planificación y ejecución de consultas

El plan de ejecución para una instrucción de consulta de Amazon Redshift específica desglosa la ejecución y el cálculo de una consulta en una secuencia discreta de operaciones de tabla y pasos que eventualmente producen un conjunto de resultados finales para la consulta. Para obtener información acerca de la planificación de consultas, consulte Procesamiento de consultas.

La siguiente tabla proporciona un resumen de los pasos que Amazon Redshift puede utilizar con objeto de desarrollar un plan de ejecución para cualquier consulta que un usuario envíe para su ejecución.

| Operadores EXPLAIN | Pasos de ejecución de la consulta | Descripción |
|---|---|---|
| SCAN: | | |
| Sequential Scan | scan | Es el paso o el operador de análisis de tabla, o el análisis de la relación de Amazon Redshift. Analiza la tabla completa de manera secuencial de principio a fin y evalúa las restricciones de la consulta para cada una de las filas (Filtro) si se especifica con la cláusula WHERE. También se utiliza para ejecutar las instrucciones INSERT, UPDATE y DELETE. |

| Operadores EXPLAIN | Pasos de ejecución de la consulta | Descripción |
|---|---|---|
| JOINS: Amazon Redshift utiliza diferentes operadores de combinación en función del diseño físico de las tablas que se combinarán, de la ubicación de los datos que se necesitan para la combinación y de los atributos específicos de la propia consulta. Análisis de subconsulta -- El anexo y análisis de subconsultas se utilizan para ejecutar consultas UNION. | | |
| Nested Loop | nloop | Combinación menos óptima; se utiliza principalmente para las combinaciones cruzadas (productos cartesianos; sin una condición de combinación) y algunas combinaciones de desigualdades. |
| Hash Join | hjoin | También se utiliza para combinaciones internas y combinaciones externas de izquierda y derecha; por lo general, es más rápida que una combinación de bucle anidado. La combinación hash lee la tabla hash externa, aplica la función hash a la columna de combinación y encuentra coincidencias en la tabla hash interna. El paso puede verterse en el disco. (La entrada interna de hjoin es un paso hash que puede estar basado en el disco). |
| Merge Join | mjoin | También se utiliza para combinaciones internas y externas (para tablas de combinación que están distribuidas y ordenadas en las columnas de combinación). Por lo general, es el algoritmo de combinación de Amazon Redshift más rápido, sin incluir otras consideraciones en torno al costo. |
| AGGREGATION: operadores y pasos que se utilizan en las consultas que implican funciones de agregación y operaciones GROUP BY. | | |

| Operadores EXPLAIN | Pasos de ejecución de la consulta | Descripción |
|---|---|---|
| Aggregate | aggr | Operador/paso para las funciones de agregación escalar. |
| HashAggregate | aggr | Operador/paso para las funciones de agregación agrupadas. Puede funcionar desde el disco en virtud de la tabla hash que se vierte en el disco. |
| GroupAggregate | aggr | En ocasiones, es el operador elegido para consultas de agrupación en conjunto si la configuración de Amazon Redshift del ajuste force_hash_grouping está desactivada. |
| SORT: operadores y pasos que se utilizan cuando las consultas tienen que ordenar o fusionar conjuntos de resultados. | | |
| Sort | sort | Sort realiza la ordenación especificada por la cláusula ORDER BY y, también, otras operaciones como UNION y combinaciones. Puede funcionar desde el disco. |
| Merge | merge | Produce los resultados ordenados finales de una consulta en función de los resultados ordenados intermedios que se derivan de las operaciones realizadas en paralelo. |
| Operaciones EXCEPT, INTERSECT y UNION: | | |
| SetOp Except [Distinct] | hjoin | Se usa para consultas EXCEPT. Puede funcionar desde el disco dado que el hash de entrada puede estar basado en el disco. |
| Hash Intersect [Distinct] | hjoin | Se usa para consultas INTERSECT. Puede funcionar desde el disco dado que el hash de entrada puede estar basado en el disco. |

| Operadores EXPLAIN | Pasos de ejecución de la consulta | Descripción |
|---|---|---|
| Append [All \|Distinct] | save | Anexo utilizado con el análisis de subconsulta para implementar las consultas UNION y UNION ALL. Puede funcionar desde el disco en virtud del comando "save". |
| Otros: | | |
| Hash | hash | Se utiliza para combinaciones internas y combinaciones externas de izquierda y derecha (proporciona una entrada a la combinación hash). El operador Hash crea la tabla hash para la tabla interna de una combinación. (La tabla interna es la tabla que se revisa para encontrar coincidencias y, en una combinación de dos tablas, es por lo general la más pequeña de las dos). |
| Límite | límite | Evalúa la cláusula LIMIT. |
| Materialize | save | Materializa las filas para ingresarlas en combinaciones de bucle anidado y en algunas combinaciones de fusión. Puede funcionar desde el disco. |
| -- | parse | Se usa para analizar datos de entrada de texto durante una carga. |
| -- | proyecto | Se usa para reordenar columnas y expresiones de computación, es decir, datos del proyecto. |
| Resultado | -- | Ejecuta funciones escalares que no implican el acceso a ninguna tabla. |
| -- | return | Devuelve filas al nodo principal o cliente. |

| Operadores EXPLAIN | Pasos de ejecución de la consulta | Descripción |
|---|---|---|
| Subplan | -- | Se utiliza para algunas subconsultas. |
| Unique | unique | Elimina los duplicados de las consultas SELECT DISTINCT y UNION. |
| Ventana | periodo | Calcula la adición y clasificación de funciones de ventana. Puede funcionar desde el disco. |
| Operaciones de red: | | |
| Network (Broadcast) | bcast | Broadcast también es una atributo de los operadores y pasos de Join Explain. |
| Network (Distribute) | dist | Distribuye filas en nodos de computación para el procesamiento paralelo del clúster del data warehouse. |
| Network (Send to Leader) | return | Envía los resultados de regreso al nodo principal para seguir trabajando con ellos. |
| Operaciones DML (operadores que modifican datos): | | |
| Insert (using Result) | inserción | Inserta datos. |
| Delete (Scan + Filter) | eliminar | Elimina datos. Puede funcionar desde el disco. |
| Update (Scan + Filter) | delete, insert | Implementado como eliminar e insertar. |

## Uso de EXPLAIN para RLS

Si una consulta contiene una tabla que está sujeta a políticas de seguridad de la fila (RLS), EXPLAIN muestra un nodo especial SecureScan de RLS. Amazon Redshift también registra el mismo tipo de nodo en la tabla del sistema STL_EXPLAIN. EXPLAIN no revela el predicado de RLS que se aplica a dim_tbl. El tipo de nodo SecureScan de RLS sirve como indicador de que el plan de ejecución contiene operaciones adicionales que son invisibles para el usuario actual.

En el siguiente ejemplo, se ilustra un nodo SecureScan de RLS.

```
EXPLAIN
SELECT D.cint
FROM fact_tbl F INNER JOIN dim_tbl D ON F.k_dim = D.k
WHERE F.k_dim / 10 > 0;
                              QUERY PLAN
------------------------------------------------------------------------
 XN Hash Join DS_DIST_ALL_NONE  (cost=0.08..0.25 rows=1 width=4)
   Hash Cond: ("outer".k_dim = "inner"."k")
   ->  *XN* *RLS SecureScan f  (cost=0.00..0.14 rows=2 width=4)*
         Filter: ((k_dim / 10) > 0)
   ->  XN Hash  (cost=0.07..0.07 rows=2 width=8)
         ->  XN Seq Scan on dim_tbl d  (cost=0.00..0.07 rows=2 width=8)
               Filter: (("k" / 10) > 0)
```

Para habilitar una investigación completa de los planes de consultas que están sujetos a RLS, Amazon Redshift ofrece los permisos del sistema EXPLAIN RLS. Los usuarios a los que se les haya otorgado este permiso pueden inspeccionar planes de consulta completos que también incluyan predicados de RLS.

En el siguiente ejemplo, se ilustra un análisis de secuencia adicional debajo del nodo SecureScan de RLS que también incluye el predicado de política de RLS (k_dim > 1).

```
EXPLAIN SELECT D.cint
FROM fact_tbl F INNER JOIN dim_tbl D ON F.k_dim = D.k
WHERE F.k_dim / 10 > 0;
                                  QUERY PLAN
------------------------------------------------------------------------------
 XN Hash Join DS_DIST_ALL_NONE  (cost=0.08..0.25 rows=1 width=4)
   Hash Cond: ("outer".k_dim = "inner"."k")
   *->  XN RLS SecureScan f  (cost=0.00..0.14 rows=2 width=4)
         Filter: ((k_dim / 10) > 0)*
         ->  *XN* *Seq Scan on fact_tbl rls_table  (cost=0.00..0.06 rows=5 width=8)
               Filter: (k_dim > 1)*
   ->  XN Hash  (cost=0.07..0.07 rows=2 width=8)
         ->  XN Seq Scan on dim_tbl d  (cost=0.00..0.07 rows=2 width=8)
               Filter: (("k" / 10) > 0)
```

Mientras se concede el permiso EXPLAIN RLS a un usuario, Amazon Redshift registra el plan de consulta completo, incluidos los predicados de RLS, en la tabla del sistema STL_EXPLAIN. Las consultas que se ejecutan mientras no se concede este permiso se registrarán sin los internos de

RLS. La concesión o eliminación del permiso EXPLAIN RLS no cambiará lo que Amazon Redshift haya registrado en STL_EXPLAIN para consultas anteriores.

Relaciones de Redshift protegidas por RLS de AWS Lake Formation

En el siguiente ejemplo se ilustra un nodo LF SecureScan que puede utilizar para ver las relaciones entre Lake Formation y RLS.

```
EXPLAIN
SELECT *
FROM lf_db.public.t_share
WHERE a > 1;
QUERY PLAN
------------------------------------------------------------
XN LF SecureScan t_share  (cost=0.00..0.02 rows=2 width=11)
(2 rows)
```

## Ejemplos

> **Note**
>
> Para estos ejemplos, la salida de ejemplo puede variar según la configuración de Amazon Redshift.

El siguiente ejemplo devuelve el plan de consulta para una consulta que selecciona EVENTID, EVENTNAME, VENUEID y VENUENAME de la tablas EVENT y VENUE:

```
explain
select eventid, eventname, event.venueid, venuename
from event, venue
where event.venueid = venue.venueid;
```

```
                              QUERY PLAN
------------------------------------------------------------------------
XN Hash Join DS_DIST_OUTER  (cost=2.52..58653620.93 rows=8712 width=43)
Hash Cond: ("outer".venueid = "inner".venueid)
->  XN Seq Scan on event  (cost=0.00..87.98 rows=8798 width=23)
->  XN Hash  (cost=2.02..2.02 rows=202 width=22)
->  XN Seq Scan on venue  (cost=0.00..2.02 rows=202 width=22)
```

```
(5 rows)
```

El siguiente ejemplo devuelve el plan de consulta para la misma consulta con resultados más detallados:

```
explain verbose
select eventid, eventname, event.venueid, venuename
from event, venue
where event.venueid = venue.venueid;
```

```
                                QUERY PLAN
------------------------------------------------------------------------------
{HASHJOIN
:startup_cost 2.52
:total_cost 58653620.93
:plan_rows 8712
:plan_width 43
:best_pathkeys <>
:dist_info DS_DIST_OUTER
:dist_info.dist_keys (
TARGETENTRY
{
VAR
:varno 2
:varattno 1
...

XN Hash Join DS_DIST_OUTER  (cost=2.52..58653620.93 rows=8712 width=43)
Hash Cond: ("outer".venueid = "inner".venueid)
->  XN Seq Scan on event  (cost=0.00..87.98 rows=8798 width=23)
->  XN Hash  (cost=2.02..2.02 rows=202 width=22)
->  XN Seq Scan on venue  (cost=0.00..2.02 rows=202 width=22)
(519 rows)
```

El siguiente ejemplo devuelve el plan de consulta para una instrucción CREATE TABLE AS (CTAS):

```
explain create table venue_nonulls as
select * from venue
where venueseats is not null;

QUERY PLAN
-------------------------------------------------------------
```

```
XN Seq Scan on venue  (cost=0.00..2.02 rows=187 width=45)
Filter: (venueseats IS NOT NULL)
(2 rows)
```

# FETCH

Recupera filas con un cursor. Para obtener información acerca de la declaración de un cursor, consulte DECLARE.

FETCH recupera filas en función de la posición actual dentro del cursor. Cuando se crea un cursor, se posiciona antes de la primera fila. Después de FETCH, el cursor se posiciona en la última fila recuperada. Si FETCH se ejecuta en el extremo de las filas disponibles, como tras un comando FETCH ALL, el cursor queda posicionado después de la última fila.

FORWARD 0 recupera la fila actual sin mover el cursor; es decir, recupera la fila que se buscó más recientemente. Si el cursor se posiciona antes de la primera fila o después de la última fila, no se devuelve una fila.

Cuando se recupera la primera fila de un cursor, el conjunto completo de resultados se materializa en el nodo principal, en la memoria o en el disco, si es necesario. Debido al posible impacto negativo en el rendimiento del uso de cursores con conjuntos de resultados grandes, recomendamos usar enfoques alternativos cuando sea posible. Para obtener más información, consulte Consideraciones acerca del rendimiento cuando se utilizan cursores.

Para obtener más información, consulte DECLARE, CLOSE.

## Sintaxis

```
FETCH [ NEXT | ALL | {FORWARD [ count | ALL ] } ] FROM cursor
```

## Parámetros

NEXT

Recupera la última fila. Esta es la opción predeterminada.

ALL

Recupera todas las filas restantes. (Igual que FORWARD ALL). ALL no es compatible con los clústeres de un solo nodo.

FORWARD [ count (número) | ALL ]

Recupera el siguiente count (número) de filas o todas las filas restantes. `FORWARD 0` recupera la fila actual. Para los clústeres de un nodo, el valor máximo del valor count es `1000`. FORWARD ALL no es compatible con los clústeres de un solo nodo.

cursor

Nombre del nuevo cursor.

## Ejemplo de FETCH

En el siguiente ejemplo, se declara un cursor denominado LOLLAPALOOZA para seleccionar información de ventas para el evento Lollapalooza y, luego, se recuperan filas del conjunto de resultados a través del cursor:

```
-- Begin a transaction

begin;

-- Declare a cursor

declare lollapalooza cursor for
select eventname, starttime, pricepaid/qtysold as costperticket, qtysold
from sales, event
where sales.eventid = event.eventid
and eventname='Lollapalooza';

-- Fetch the first 5 rows in the cursor lollapalooza:

fetch forward 5 from lollapalooza;

  eventname   |      starttime      | costperticket | qtysold
--------------+---------------------+---------------+---------
 Lollapalooza | 2008-05-01 19:00:00 |   92.00000000 |       3
 Lollapalooza | 2008-11-15 15:00:00 |  222.00000000 |       2
 Lollapalooza | 2008-04-17 15:00:00 |  239.00000000 |       3
 Lollapalooza | 2008-04-17 15:00:00 |  239.00000000 |       4
 Lollapalooza | 2008-04-17 15:00:00 |  239.00000000 |       1
(5 rows)

-- Fetch the next row:
```

```
fetch next from lollapalooza;

  eventname   |      starttime      | costperticket | qtysold
--------------+---------------------+---------------+---------
 Lollapalooza | 2008-10-06 14:00:00 |  114.00000000 |       2

-- Close the cursor and end the transaction:

close lollapalooza;
commit;
```

## GRANT

Define permisos de acceso para un usuario o rol.

Los permisos incluyen opciones de acceso, como poder leer datos en tablas y vistas, escribir datos, crear tablas y eliminarlas. Utilice este comando para otorgar permisos específicos a una tabla, base de datos, esquema, función, procedimiento, lenguaje o columna. Para revocar los permisos de un objeto de la base de datos, utilice el comando REVOKE.

Los permisos también incluyen las siguientes opciones de acceso al productor de recursos compartidos de datos:

- Otorgar acceso a un recurso compartido de datos a los espacios de nombres y las cuentas del consumidor.
- Otorgar permiso para modificar un recurso compartido de datos mediante la adición o eliminación de objetos del recurso compartido de datos.
- Otorgar permiso para compartir un recurso compartido de datos mediante la adición o eliminación de espacios de nombres del consumidor del recurso compartido de datos.

Las opciones de acceso para consumidores de recursos compartidos de datos son las siguientes:

- Otorgar a los usuarios acceso total a las bases de datos creadas a partir de un recurso compartido de datos o a esquemas externos que apuntan a dichas bases de datos.
- Otorgar a los usuarios permisos de nivel de objeto en las bases de datos creadas a partir de un recurso compartido de datos, tal como se hace con los objetos de bases de datos locales. Para conceder este nivel de permiso, debe utilizar la cláusula WITH PERMISSIONS al crear una base de datos a partir del recurso compartido de datos. Para obtener más información, consulte CREATE DATABASE.

Para obtener más información sobre permisos de recursos compartidos de datos, consulte Permisos que puede conceder a los recursos compartidos de datos.

También puede conceder roles para administrar los permisos de base de datos y controlar lo que los usuarios pueden hacer en relación con los datos. Al definir roles y asignarlos a los usuarios, puede limitar las acciones que estos pueden realizar, como, por ejemplo, limitar a los usuarios solo a los comandos CREATE TABLE e INSERT. Para obtener más información sobre el comando CREATE ROLE, consulte the section called “CREAR ROL”. Amazon Redshift dispone de algunos roles definidos por el sistema que también puede utilizar para conceder permisos específicos a sus usuarios. Para obtener más información, consulte the section called “Roles definidos por el sistema de Amazon Redshift”.

Solo puede usar permisos GRANT o REVOKE USAGE en un esquema externo de los usuarios de la base de datos y los grupos de usuarios que utilicen la sintaxis ON SCHEMA. Cuando use ON EXTERNAL SCHEMA con AWS Lake Formation, solo puede usar GRANT y REVOKE para conceder y revocar permisos a un rol de AWS Identity and Access Management (IAM). Para obtener una lista de los permisos, consulte la sintaxis.

Para los procedimientos almacenados, el único permiso que puede concederse es EXECUTE.

No se puede ejecutar GRANT (en un recurso externo) en un bloque de transacción (BEGIN … END). Para obtener más información acerca de las transacciones, consulte Aislamiento serializable.

Para ver qué permisos se han concedido a los usuarios para una base de datos, utilice HAS_DATABASE_PRIVILEGE. Para ver qué permisos se han concedido a los usuarios para un esquema, utilice HAS_SCHEMA_PRIVILEGE. Para ver qué permisos se han concedido a los usuarios para una tabla, utilice HAS_TABLE_PRIVILEGE.

## Sintaxis

```
GRANT { { SELECT | INSERT | UPDATE | DELETE | DROP | REFERENCES | ALTER | TRUNCATE }
 [,...] | ALL [ PRIVILEGES ] }
    ON { [ TABLE ] table_name [, ...] | ALL TABLES IN SCHEMA schema_name [, ...] }
    TO { username [ WITH GRANT OPTION ] | ROLE role_name | GROUP group_name | PUBLIC }
 [, ...]

GRANT { { CREATE | USAGE | TEMPORARY | TEMP | ALTER } [,...] | ALL [ PRIVILEGES ] }
    ON DATABASE db_name [, ...]
    TO { username [ WITH GRANT OPTION ] | ROLE role_name | GROUP group_name | PUBLIC }
 [, ...]
```

```
GRANT { { CREATE | USAGE | ALTER | DROP } [,...] | ALL [ PRIVILEGES ] }
    ON SCHEMA schema_name [, ...]
    TO { username [ WITH GRANT OPTION ] | ROLE role_name | GROUP group_name | PUBLIC }
 [, ...]

GRANT { EXECUTE | ALL [ PRIVILEGES ] }
    ON { FUNCTION function_name ( [ [ argname ] argtype [, ...] ] ) [, ...] | ALL
 FUNCTIONS IN SCHEMA schema_name [, ...] }
    TO { username [ WITH GRANT OPTION ] | ROLE role_name | GROUP group_name | PUBLIC }
 [, ...]

GRANT { EXECUTE | ALL [ PRIVILEGES ] }
    ON { PROCEDURE procedure_name ( [ [ argname ] argtype [, ...] ] ) [, ...] | ALL
 PROCEDURES IN SCHEMA schema_name [, ...] }
    TO { username [ WITH GRANT OPTION ] | ROLE role_name | GROUP group_name | PUBLIC }
 [, ...]

GRANT USAGE
    ON LANGUAGE language_name [, ...]
    TO { username [ WITH GRANT OPTION ] | ROLE role_name | GROUP group_name | PUBLIC }
 [, ...]

GRANT { { ALTER | DROP} [,...] | ALL [ PRIVILEGES ] }
    ON COPY JOB job_name [,...]
    TO { username [ WITH GRANT OPTION ] | ROLE role_name | GROUP group_name | PUBLIC }
 [, ...]
```

Concesión de permisos de nivel de columna para tablas

A continuación, se muestra la sintaxis de los permisos de nivel de columna en tablas y vistas de Amazon Redshift.

```
GRANT { { SELECT | UPDATE } ( column_name [, ...] ) [, ...] | ALL [ PRIVILEGES ]
 ( column_name [,...] ) }
     ON { [ TABLE ] table_name [, ...] }

     TO { username | ROLE role_name | GROUP group_name | PUBLIC } [, ...]
```

Concesión de permisos ASSUMEROLE

A continuación, se muestra la sintaxis de los permisos ASSUMEROLE concedidos a usuarios y grupos con un rol especificado. Para empezar a utilizar el privilegio ASSUMEROLE, consulte Notas de uso para conceder el permiso ASSUMEROLE.

```
GRANT ASSUMEROLE
       ON { 'iam_role' [, ...] | default | ALL }
       TO { username | ROLE role_name | GROUP group_name | PUBLIC } [, ...]
       FOR { ALL | COPY | UNLOAD | EXTERNAL FUNCTION | CREATE MODEL } [, ...]
```

Concesión de permisos para la integración de Redshift Spectrum con Lake Formation

A continuación, se muestra la sintaxis para la integración de Redshift Spectrum con Lake Formation.

```
GRANT { SELECT | ALL [ PRIVILEGES ] } ( column_list )
    ON EXTERNAL TABLE schema_name.table_name
    TO { IAM_ROLE iam_role } [, ...] [ WITH GRANT OPTION ]

GRANT { { SELECT | ALTER | DROP | DELETE | INSERT }  [, ...] | ALL [ PRIVILEGES ] }
    ON EXTERNAL TABLE schema_name.table_name [, ...]
    TO { { IAM_ROLE iam_role } [, ...] | PUBLIC } [ WITH GRANT OPTION ]

GRANT { { CREATE | ALTER | DROP }  [, ...] | ALL [ PRIVILEGES ] }
    ON EXTERNAL SCHEMA schema_name [, ...]
    TO { IAM_ROLE iam_role } [, ...] [ WITH GRANT OPTION ]
```

Concesión de permisos de recurso compartido de datos

Permisos de recursos compartidos de datos del lado del productor

A continuación, se muestra la sintaxis para utilizar GRANT a fin de conceder permisos ALTER o SHARE a un usuario o rol. El usuario puede modificar el recurso compartido de datos con el permiso ALTER o conceder el uso a un consumidor con el permiso SHARE. ALTER y SHARE son los únicos permisos que se pueden conceder en recursos compartidos de datos a usuarios y a grupos de usuarios.

```
GRANT { ALTER | SHARE } ON DATASHARE datashare_name
    TO { username [ WITH GRANT OPTION ] | ROLE role_name | GROUP group_name | PUBLIC }
 [, ...]
```

A continuación, se muestra la sintaxis para utilizar GRANT en los permisos de uso de recurso compartido de datos en Amazon Redshift. Con el permiso USAGE se concede acceso a un recurso compartido de datos a un consumidor. No se puede conceder este permiso a usuarios o grupos de usuarios. Este permiso tampoco admite WITH GRANT OPTION para la instrucción GRANT. Solo los usuarios o los grupos de usuarios a los que se haya concedido el permiso SHARE con anterioridad para el recurso compartido de datos pueden ejecutar este tipo de instrucción GRANT.

```
GRANT USAGE
    ON DATASHARE datashare_name
    TO NAMESPACE 'namespaceGUID' | ACCOUNT 'accountnumber' [ VIA DATA CATALOG ]
```

A continuación, se ofrece un ejemplo de cómo conceder el uso de un recurso compartido de datos a una cuenta de Lake Formation.

```
GRANT USAGE ON DATASHARE salesshare TO ACCOUNT '123456789012' VIA DATA CATALOG;
```

Permisos de recursos compartidos de datos del lado del consumidor

A continuación, se muestra la sintaxis para los permisos de uso compartido de datos GRANT en una base de datos específica o un esquema específico creados a partir de un datashare.

Los demás permisos necesarios para que los consumidores accedan a una base de datos creada a partir de un recurso compartido de datos varían en función de si el comando CREATE DATABASE utilizado para crear la base de datos a partir del recurso compartido de datos ha utilizado o no la cláusula WITH PERMISSIONS. Para obtener más información acerca del comando CREATE DATABASE y la cláusula WITH PERMISSIONS, consulte CREATE DATABASE.

Bases de datos creadas sin usar la cláusula WITH PERMISSIONS

Al conceder USAGE a una base de datos creada a partir de un recurso compartido de datos sin la cláusula WITH PERMISSIONS, no es necesario conceder permisos por separado a los objetos de la base de datos compartida. Las entidades a las que se concede el uso en bases de datos creadas a partir de recursos compartidos de datos sin la cláusula WITH PERMISSIONS tienen acceso automáticamente a todos los objetos de la base de datos.

Bases de datos creadas con la cláusula WITH PERMISSIONS

Al conceder USAGE a una base de datos en la que se ha creado la base de datos compartida a partir de un recurso compartido de datos con la cláusula WITH PERMISSIONS, a las identidades

del lado del consumidor se les deben seguir concediendo los permisos pertinentes para los objetos de la base de datos compartida para acceder a ellas, del mismo modo que se concederían permisos a los objetos de la base de datos local. Para conceder permisos a los objetos de una base de datos creada a partir de un recurso compartido de datos, utilice la sintaxis de tres partes `database_name.schema_name.object_name`. Para conceder permisos a los objetos en un esquema externo que apunte a un esquema compartido dentro de la base de datos compartida, utilice la sintaxis de dos partes `schema_name.object_name`.

```
GRANT USAGE ON { DATABASE shared_database_name [, ...] | SCHEMA shared_schema}
    TO { username | ROLE role_name | GROUP group_name | PUBLIC } [, ...]
```

Concesión de permisos acotados

Los permisos limitados le permiten conceder permisos a un usuario o rol en todos los objetos de un tipo dentro de una base de datos o un esquema. Los usuarios y roles con permisos limitados tienen los permisos especificados en todos los objetos actuales y futuros de la base de datos o del esquema.

Puede ver el alcance de los permisos limitados en el nivel de base de datos en SVV_DATABASE_PRIVILEGES. Puede ver el alcance de los permisos limitados en el nivel de esquema en SVV_SCHEMA_PRIVILEGES.

Para obtener más información sobre los permisos acotados, consulte Permisos acotados.

A continuación, se muestra la sintaxis para conceder permisos acotados a usuarios y roles.

```
GRANT { CREATE | USAGE | ALTER | DROP } [,...] | ALL [ PRIVILEGES ] }
FOR SCHEMAS IN
DATABASE db_name
TO { username [ WITH GRANT OPTION ] | ROLE role_name } [, ...]

GRANT
{ { SELECT | INSERT | UPDATE | DELETE | DROP | ALTER | TRUNCATE | REFERENCES }
 [, ...] } | ALL [PRIVILEGES] } }
FOR TABLES IN
{SCHEMA schema_name [DATABASE db_name ] | DATABASE db_name }
TO { username [ WITH GRANT OPTION ] | ROLE role_name} [, ...]

GRANT { EXECUTE | ALL [ PRIVILEGES ] }
FOR FUNCTIONS IN
{SCHEMA schema_name [DATABASE db_name ] | DATABASE db_name }
```

```
TO { username [ WITH GRANT OPTION ] | ROLE role_name | } [, ...]

GRANT { EXECUTE | ALL [ PRIVILEGES ] }
FOR PROCEDURES IN
{SCHEMA schema_name [DATABASE db_name ] | DATABASE db_name }
TO { username [ WITH GRANT OPTION ] | ROLE role_name | } [, ...]

GRANT USAGE
FOR LANGUAGES IN
{DATABASE db_name}
TO { username [ WITH GRANT OPTION ] | ROLE role_name } [, ...]

GRANT { { CREATE | ALTER | DROP} [,...] | ALL [ PRIVILEGES ] }
FOR COPY JOBS
IN DATABASE db_name
TO { username [ WITH GRANT OPTION ] | ROLE role_name } [, ...]
```

Tenga en cuenta que los permisos limitados no distinguen entre los permisos de las funciones y los de los procedimientos. Por ejemplo, la siguiente instrucción concede a bob el permiso EXECUTE tanto para las funciones como para los procedimientos en el esquema Sales_schema.

```
GRANT EXECUTE FOR FUNCTIONS IN SCHEMA Sales_schema TO bob;
```

Concesión de permisos de machine learning

A continuación, se muestra la sintaxis para los permisos de modelos de machine learning en Amazon Redshift.

```
GRANT CREATE MODEL
    TO { username [ WITH GRANT OPTION ] | ROLE role_name | GROUP group_name | PUBLIC }
 [, ...]

GRANT { EXECUTE | ALL [ PRIVILEGES ] }
    ON MODEL model_name [, ...]

    TO { username [ WITH GRANT OPTION ] | ROLE role_name | GROUP group_name | PUBLIC }
 [, ...]
```

Concesión de permisos de rol

La siguiente es la sintaxis para conceder roles en Amazon Redshift.

```
GRANT { ROLE role_name } [, ...] TO { { user_name [ WITH ADMIN OPTION ] } |
 ROLE role_name }[, ...]
```

A continuación, se muestra la sintaxis para conceder permisos de sistema a los roles en Amazon Redshift. Tenga en cuenta que solo puede conceder permisos a los roles, no a los usuarios.

```
GRANT
  {
    { CREATE USER | DROP USER | ALTER USER |
    CREATE SCHEMA | DROP SCHEMA |
    ALTER DEFAULT PRIVILEGES |
    ACCESS CATALOG | ACCESS SYSTEM TABLE
    CREATE TABLE | DROP TABLE | ALTER TABLE |
    CREATE OR REPLACE FUNCTION | CREATE OR REPLACE EXTERNAL FUNCTION |
    DROP FUNCTION |
    CREATE OR REPLACE PROCEDURE | DROP PROCEDURE |
    CREATE OR REPLACE VIEW | DROP VIEW |
    CREATE MODEL | DROP MODEL |
    CREATE DATASHARE | ALTER DATASHARE | DROP DATASHARE |
    CREATE LIBRARY | DROP LIBRARY |
    CREATE ROLE | DROP ROLE |
    TRUNCATE TABLE
    VACUUM | ANALYZE | CANCEL |
    IGNORE RLS | EXPLAIN RLS |
    EXPLAIN MASKING }[, ...]
  }
  | { ALL [ PRIVILEGES ] }
TO ROLE role_name [, ...]
```

Concesión de permisos EXPLAIN para filtros de políticas de seguridad de nivel de fila

A continuación, se muestra la sintaxis para conceder permisos para explicar los filtros de la política de seguridad de la fila de una consulta en el plan EXPLAIN. Puede revocar el privilegio mediante la instrucción REVOKE.

```
GRANT EXPLAIN RLS TO ROLE rolename
```

Con la siguiente sintaxis, se puede conceder permisos para omitir las políticas de seguridad de la fila para una consulta.

```
GRANT IGNORE RLS TO ROLE rolename
```

Concesión de permisos para tablas de búsqueda de RLS a un objeto de política

Con la siguiente sintaxis, se puede conceder permisos para omitir la política de seguridad de la fila especificada.

```
GRANT SELECT ON [ TABLE ] table_name [, ...]
TO RLS POLICY policy_name [, ...]
```

## Parámetros

SELECT

Concede permiso para seleccionar datos de una tabla o vista mediante una instrucción SELECT. También se requiere el permiso SELECT para hacer referencia a los valores de columna existentes para las operaciones UPDATE o DELETE.

INSERT

Concede el permiso para cargar datos en una tabla mediante una instrucción INSERT o COPY.

UPDATE

Concede el privilegio para actualizar una columna de tabla mediante una instrucción UPDATE. Las operaciones UPDATE también requieren el permiso SELECT, ya que deben hacer referencia a las columnas de tabla para determinar cuáles son las filas que se deben actualizar o para calcular nuevos valores para las columnas.

DELETE

Concede el permiso para eliminar una fila de datos de una tabla. Las operaciones DELETE también requieren el permiso SELECT, ya que deben hacer referencia a las columnas de tabla para determinar cuáles son las filas que se deben eliminar.

DROP

Según el objeto de base de datos, concede los siguientes permisos al usuario o al rol:

- En el caso de las tablas, DROP concede permiso para eliminar una tabla o una vista. Para obtener más información, consulte DROP TABLE.
- En el caso de las bases de datos, DROP concede permiso para eliminar una base de datos. Para obtener más información, consulte DROP DATABASE.
- En el caso de los esquemas, DROP concede permiso para eliminar un esquema. Para obtener más información, consulte DROP SCHEMA.

REFERENCES

Concede el permiso para crear una restricción de clave externa. Es necesario conceder este permiso en la tabla a la que se hace referencia y en la tabla que hace la referencia; de lo contrario, el usuario no podrá crear la restricción.

ALTER

Según el objeto de base de datos, concede los siguientes permisos al usuario o al grupo de usuarios:

- En el caso de las tablas, ALTER concede permiso para modificar una tabla o una vista. Para obtener más información, consulte ALTER TABLE.
- En el caso de las bases de datos, ALTER concede permiso para modificar una base de datos. Para obtener más información, consulte ALTER DATABASE.
- En el caso de los esquemas, ALTER concede permiso para modificar un esquema. Para obtener más información, consulte ALTER SCHEMA.
- En el caso de las tablas externas, ALTER concede permiso para alterar una tabla en un AWS Glue Data Catalog que esté habilitado para Lake Formation. Este permiso solo se aplica cuando se utiliza Lake Formation.

TRUNCATE

Concede permiso para truncar una tabla. Sin este permiso, solo el propietario de una tabla o un superusuario pueden truncarla. Para obtener más información sobre el comando TRUNCATE, consulte the section called “TRUNCATE”.

ALL [ PRIVILEGES ]

Concede todos los permisos disponibles a la vez al usuario o al rol especificado. La palabra clave PRIVILEGES es opcional.

GRANT ALL ON SCHEMA no concede permisos CREATE para esquemas externos.

Puede conceder el permiso ALL a una tabla en un AWS Glue Data Catalog que está habilitado para Lake Formation. En este caso, los permisos individuales (como SELECT, ALTER, etc.) se registran en el catálogo de datos.

> **Note**
>
> Amazon Redshift no admite los permisos RULE y TRIGGER. Para obtener más información, consulte Características no compatibles de PostgreSQL.

ASSUMEROLE

Concede permiso para ejecutar los comandos COPY, UNLOAD, EXTERNAL FUNCTION y CREATE MODEL a usuarios, roles o grupos con un rol especificado. El usuario, el rol o el grupo asume ese rol cuando se ejecuta el comando especificado. Para empezar a utilizar el permiso ASSUMEROLE, consulte Notas de uso para conceder el permiso ASSUMEROLE.

ON [ TABLE ] table_name

Concede los permisos especificados en una tabla o vista. La palabra clave TABLE es opcional. Puede enumerar varias tablas y vistas en una instrucción.

ON ALL TABLES IN SCHEMA schema_name (nombre_de_esquema)

Concede los permisos especificados en todas las tablas y vistas en el esquema al que se hace referencia.

( column_name [,...] ) ON TABLE table_name

Concede los permisos especificados a usuarios, grupos o PUBLIC en las columnas especificadas de la tabla o la vista de Amazon Redshift.

( column_list ) ON EXTERNAL TABLE schema_name.table_name

Concede los permisos especificados a un rol de IAM en las columnas especificadas de la tabla de Lake Formation en el esquema al que se hace referencia.

ON EXTERNAL TABLE schema_name.table_name

Concede los permisos especificados a un rol de IAM en las tablas de Lake Formation especificadas en el esquema al que se hace referencia.

ON EXTERNAL SCHEMA schema_name

Concede los permisos especificados a un rol de IAM en el esquema al que se hace referencia.

ON iam_role

Concede los permisos especificados a un rol de IAM.

TO username

Indica el usuario que recibe los permisos.

TO IAM_ROLE iam_role

Indica el rol de IAM que recibe los permisos.

WITH GRANT OPTION

Indica que el usuario que recibe los permisos puede, a su vez, conceder los mismos permisos a otros usuarios. No se puede conceder WITH GRANT OPTION a un grupo o a PUBLIC.

ROLE role_name

Concede el permiso a un rol.

GROUP group_name

Concede los permisos para a un grupo de usuarios. Puede ser una lista separada por comas para especificar varios grupos de usuarios.

PUBLIC

Concede los permisos especificados a todos los usuarios, incluidos los creados posteriormente. PUBLIC representa un grupo que siempre incluye a todos los usuarios. Los permisos de un usuario individual constan de la suma de permisos concedidos a PUBLIC, los permisos concedidos a cualquier grupo al que pertenezca el usuario y los permisos concedidos al usuario de manera individual.

La concesión de PUBLIC a una tabla externa de Lake Formation da lugar a la concesión del permiso al grupo everyone de Lake Formation.

CREATE

Según el objeto de base de datos, concede los siguientes permisos al usuario o al grupo de usuarios:

- Para las bases de datos, CREATE permite que los usuarios creen esquemas dentro de la base de datos.
- Para los esquemas, CREATE permite que los usuarios creen objetos dentro de un esquema. Para cambiar el nombre de un objeto, el usuario debe tener el permiso CREATE y ser propietario del objeto cuyo nombre va a cambiarse.
- CREATE ON SCHEMA no se admite para esquemas externos de Amazon Redshift Spectrum. Para conceder permisos para usar tablas externas en un esquema externo, conceda el privilegio USAGE ON SCHEMA a los usuarios que necesitan tener acceso. Solo el propietario de un esquema externo o un superusuario pueden crear tablas externas en el esquema externo. Para transferir la propiedad de un esquema externo, use ALTER SCHEMA para cambiar el propietario.

TEMPORARY | TEMP

Concede el permiso para crear tablas temporales en la base de datos especificada. Para ejecutar consultas de Amazon Redshift Spectrum, el usuario de la base de datos debe tener permiso para crear tablas temporales en la base de datos.

> **Note**
>
> Por defecto, los usuarios reciben permisos para crear tablas temporales con su membresía automática en el grupo PUBLIC. Para quitar el permiso para que cualquier usuario pueda crear tablas temporales, revoque el permiso TEMP del grupo PUBLIC. A continuación, conceda explícitamente el permiso para crear tablas temporales a usuarios o grupos de usuarios específicos.

ON DATABASE db_name

Concede los permisos especificados en una base de datos.

USAGE

Concede el permiso USAGE en un esquema específico, lo que permite que los usuarios obtengan acceso a los objetos de ese esquema. Las acciones específicas en estos objetos deben concederse por separado (por ejemplo, permisos SELECT o UPDATE en las tablas) para los esquemas locales de Amazon Redshift. De forma predeterminada, todos los usuarios tienen permisos CREATE y USAGE en el esquema PUBLIC.

Al conceder USAGE a esquemas externos mediante la sintaxis ON SCHEMA, no es necesario conceder acciones por separado a los objetos del esquema externo. Los permisos de catálogo correspondientes controlan los permisos detallados de los objetos de esquema externos.

ON SCHEMA schema_name

Concede los permisos especificados en un esquema.

GRANT CREATE ON SCHEMA y el permiso CREATE en GRANT ALL ON SCHEMA no se admiten para esquemas externos de Amazon Redshift Spectrum. Para conceder permisos para usar tablas externas en un esquema externo, conceda el privilegio USAGE ON SCHEMA a los usuarios que necesitan tener acceso. Solo el propietario de un esquema externo o un superusuario pueden crear tablas externas en el esquema externo. Para transferir la propiedad de un esquema externo, use ALTER SCHEMA para cambiar el propietario.

EXECUTE ON ALL FUNCTIONS IN SCHEMA schema_name

Concede los permisos especificados en todas las funciones en el esquema al que se hace referencia.

Amazon Redshift no admite las instrucciones GRANT ni REVOKE para las entradas integradas de pg_proc definidas en el espacio de nombres pg_catalog.

EXECUTE ON PROCEDURE procedure_name

Concede el permiso EXECUTE en un procedimiento almacenado específico. Debido a que los nombres de procedimientos almacenados no se pueden sobrecargar, debe incluir la lista de argumentos para el procedimiento. Para obtener más información, consulte Nomenclatura de los procedimientos almacenados.

EXECUTE ON ALL PROCEDURES IN SCHEMA schema_name

Concede los permisos especificados en todos los procedimientos almacenados en el esquema al que se hace referencia.

USAGE ON LANGUAGE language_name

Concede el permiso USAGE en un lenguaje.

Se requiere el permiso USAGE ON LANGUAGE para crear funciones definidas por el usuario (UDF) mediante la ejecución del comando CREATE FUNCTION. Para obtener más información, consulte Seguridad y permisos de UDF.

Se requiere el permiso USAGE ON LANGUAGE para crear procedimientos almacenados mediante la ejecución del comando CREATE PROCEDURE. Para obtener más información, consulte Seguridad y privilegios para procedimientos almacenados .

Para las UDF de Python, use `plpythonu`. Para las UDF de SQL, use `sql`. Para procedimientos almacenados, utilice `plpgsql`.

ON COPY JOB job_name

Concede los permisos especificados en un trabajo de copia.

FOR { ALL | COPY | UNLOAD | EXTERNAL FUNCTION | CREATE MODEL } [, ...]

Especifica el comando SQL para el que se concede el permiso. Puede especificar ALL para conceder el permiso en las instrucciones COPY, UNLOAD, EXTERNAL FUNCTION y CREATE MODEL. Esta cláusula se aplica solo a la concesión del permiso ASSUMEROLE.

ALTER

Concede el permiso ALTER a los usuarios para agregar o quitar objetos de un recurso compartido de datos o para establecer la propiedad PUBLICACCESSIBLE. Para obtener más información, consulte ALTER DATASHARE.

SHARE

Concede permisos a usuarios y grupos de usuarios para agregar consumidores de datos a un recurso compartido de datos. Este permiso es necesario para permitir que un consumidor determinado (cuenta o espacio de nombres) acceda al recurso compartido de datos desde sus clústeres. El consumidor puede ser la misma cuenta de AWS o una diferente, con el mismo espacio de nombres del clúster o con uno diferente, como lo especifica un identificador global único (GUID).

ON DATASHARE datashare_name

Concede los permisos especificados en el recurso compartido de datos al que se hace referencia. Para obtener información sobre el grado de detalle del control de acceso de los consumidores, consulte Uso compartido de datos en niveles diferentes en Amazon Redshift.

USAGE

Cuando USAGE se concede a una cuenta consumidora o un espacio de nombres de la misma cuenta, estos pueden acceder al datashare y a los objetos del datashare para lectura solamente.

TO NAMESPACE 'clusternamespace GUID'

Indica un espacio de nombres en la misma cuenta donde los consumidores pueden recibir los permisos especificados para el recurso compartido de datos. Los espacios de nombres utilizan un GUID alfanumérico de 128 bits.

TO ACCOUNT 'número_de_cuenta' [ VIA DATA CATALOG ]

Indica el número de otra cuenta cuyos consumidores pueden recibir los permisos especificados para el recurso compartido de datos. Si especifica “VIA DATA CATALOG”, se indica que concede el uso del recurso compartido de datos a una cuenta de Lake Formation. La omisión de este parámetro significa que concede el uso a una cuenta que es propietaria del clúster.

ON DATABASE shared_database_name> [, ...]

Concede los permisos de uso especificados en la base de datos concreta que se crea en el recurso compartido de datos indicado.

ON SCHEMA shared_schema

Concede los permisos especificados en el esquema concreto que se crea en el recurso compartido de datos indicado.

FOR { SCHEMAS | TABLES | FUNCTIONS | PROCEDURES | LANGUAGES | COPY JOBS} IN

Especifica los objetos de la base de datos a los que se va a conceder permiso. Los parámetros después de IN definen el alcance del permiso concedido.

CREATE MODEL

Concede el permiso CREATE MODEL a usuarios o grupos de usuarios específicos.

ON MODEL model_name

Concede el permiso EXECUTE en un modelo específico.

ACCESS CATALOG

Otorga el permiso para ver los metadatos relevantes de los objetos a los que tiene acceso el rol.

{ role } [, …]

Rol que se debe conceder a otro rol, usuario o PUBLIC.

PUBLIC representa un grupo que siempre incluye a todos los usuarios. Los permisos de un usuario individual constan de la suma de permisos concedidos a PUBLIC, los permisos concedidos a cualquier grupo al que pertenezca el usuario y los permisos concedidos al usuario de manera individual.

TO { { user_name [ WITH ADMIN OPTION ] } | role }[, …]

Concede el rol especificado a un usuario especificado con WITH ADMIN OPTION, otro rol o PUBLIC.

La cláusula WITH ADMIN OPTION proporciona las opciones de administración de todos los roles concedidos a todos los beneficiarios de las concesiones.

EXPLAIN RLS TO ROLE nombre_de_rol

Concede a un rol el permiso de explicar los filtros de política de seguridad en el nivel de fila de una consulta en el plan EXPLAIN.

IGNORE RLS TO ROLE nombre_de_rol

Concede a un rol el permiso de eludir las políticas de seguridad de nivel de fila para una consulta.

## Notas de uso

Para obtener más información acerca de las notas de uso de GRANT, consulte the section called “Notas de uso”.

## Ejemplos

Para ver ejemplos de cómo utilizar GRANT, consulte the section called “Ejemplos”.

## Notas de uso

Para conceder privilegios en un objeto, debe cumplir con uno de los siguientes criterios:

- Ser el propietario del objeto.
- Ser un superusuario.
- Tener un privilegio concedido para ese objeto y privilegio.

Por ejemplo, el siguiente comando permite al usuario HR realizar comandos SELECT en la tabla de empleados y conceder y revocar el mismo privilegio para otros usuarios:

```
grant select on table employees to HR with grant option;
```

HR no puede conceder privilegios para ninguna operación que no sea SELECT o en ninguna tabla que no sea la de empleados.

Como otro ejemplo, el siguiente comando permite al usuario HR ejecutar comandos ALTER en la tabla de empleados y conceder y revocar el mismo privilegio para otros usuarios.

```
grant ALTER on table employees to HR with grant option;
```

HR no puede conceder privilegios para ninguna otra operación que no sea ALTER ni sobre ninguna otra tabla que no sea employees.

Tener privilegios concedidos de una vista no implica tener privilegios en las tablas subyacentes. De manera similar, tener privilegios concedidos de un esquema no implica tener privilegios en las tablas del esquema. En su lugar, debe conceder acceso a las tablas subyacentes de manera explícita.

Para otorgar privilegios a una tabla AWS Lake Formation, el rol IAM asociado con el esquema externo de la tabla debe tener permiso para otorgar privilegios en la tabla externa. El siguiente

ejemplo crea un esquema externo con un rol IAM `myGrantor` asociado. El rol IAM `myGrantor` tiene el permiso para otorgar permisos a otros. El comando GRANT utiliza los permisos del rol `myGrantor` IAM que está asociado al esquema externo para otorgar permisos al rol IAM `myGrantee`.

```
create external schema mySchema
from data catalog
database 'spectrum_db'
iam_role 'arn:aws:iam::123456789012:role/myGrantor'
create external database if not exists;
```

```
grant select
on external table mySchema.mytable
to iam_role 'arn:aws:iam::123456789012:role/myGrantee';
```

Si concede privilegios GRANT ALL a un rol de IAM, los privilegios individuales se conceden en el catálogo de datos relacionado y habilitado para Lake Formation. Por ejemplo, el siguiente permiso GRANT ALL da lugar a los privilegios individuales concedidos (SELECT, ALTER, DROP, DELETE e INSERT) que se muestran en la consola de Lake Formation.

```
grant all
on external table mySchema.mytable
to iam_role 'arn:aws:iam::123456789012:role/myGrantee';
```

Los superusuarios pueden obtener acceso a todos los objetos independientemente de los comandos GRANT y REVOKE que establecen privilegios de objeto.

Notas de uso para el control de acceso de nivel de columna

Las siguientes notas de uso se aplican a los privilegios del nivel de la columna en tablas y vistas de Amazon Redshift. Estas notas describen tablas; las mismas notas se aplican a las vistas a menos que anotemos explícitamente una excepción.

- En una tabla de Amazon Redshift, solo se pueden conceder los privilegios SELECT y UPDATE en el nivel de la columna. En una vista de Amazon Redshift, solo se puede conceder el privilegio SELECT en el nivel de la columna.
- La palabra clave ALL es sinónimo de una combinación de los privilegios SELECT y UPDATE cuando se utiliza en el contexto de GRANT en el nivel de columna en una tabla.
- Si no tiene el privilegio SELECT en todas las columnas de una tabla, al realizar una operación SELECT * solo se devolverán las columnas a las que tiene acceso. Cuando se utiliza una vista,

la operación SELECT * intenta acceder a todas las columnas de la vista. Si no tiene permiso para acceder a todas las columnas, estas consultas producen un error de permiso denegado.

- SELECT * no se expande solo a columnas accesibles en los siguientes casos:
  - No puede crear una vista normal solo con columnas accesibles mediante SELECT *.
  - No puede crear una vista materializada solo con columnas accesibles mediante SELECT *.
- Si tiene privilegios SELECT o UPDATE en una tabla o vista y agrega una columna, seguirá teniendo los mismos privilegios en la tabla o vista y, por tanto, en todas sus columnas.
- Sólo el propietario de una tabla o un superusuario pueden conceder privilegios de nivel de columna.
- No se admite la cláusula WITH GRANT OPTION para privilegios de nivel de columna.
- No puede mantener el mismo privilegio tanto en el nivel de tabla como en el nivel de columna. Por ejemplo, el usuario `data_scientist` no puede tener tanto el privilegio SELECT en la tabla `employee` como el privilegio SELECT en la columna `employee.department`. Tenga en cuenta los siguientes resultados al conceder el mismo privilegio a una tabla y a una columna dentro de la tabla:
  - Si un usuario tiene un privilegio de nivel de tabla en una tabla, conceder el mismo privilegio en el nivel de columna no tiene ningún efecto.
  - Si un usuario tiene un privilegio de nivel de tabla en una tabla, al revocar el mismo privilegio para una o varias columnas de la tabla se devuelve un error. En su lugar, revoque el privilegio en el nivel de tabla.
  - Si un usuario tiene un privilegio de nivel de columna, conceder el mismo privilegio en el nivel de tabla devuelve un error.
  - Si un usuario tiene un privilegio de nivel de columna, al revocar el mismo privilegio en el nivel de tabla se revocarán los privilegios de columna y tabla para todas las columnas de la tabla.
- No se pueden conceder privilegios de nivel de columna en vistas de enlace en tiempo de ejecución.
- Debe tener el privilegio SELECT de nivel de tabla en las tablas base para crear una vista materializada. Incluso si tiene privilegios de nivel de columna en columnas específicas, no puede crear una vista materializada solo en esas columnas. No obstante, puede conceder el privilegio SELECT a las columnas de una vista materializada, similar a las vistas normales.
- Para buscar concesiones de privilegios de nivel de columna, utilice la vista PG_ATTRIBUTE_INFO .

### Notas de uso para conceder el permiso ASSUMEROLE

Las siguientes notas de uso se aplican cuando se concede el permiso ASSUMEROLE en Amazon Redshift.

El permiso ASSUMEROLE se utiliza para controlar los permisos de acceso de rol de IAM para usuarios, roles y grupos de base de datos en comandos como COPY, UNLOAD, EXTERNAL FUNCTION o CREATE MODEL. Después de conceder el permiso ASSUMEROLE a un usuario, un rol o un grupo para un rol de IAM, el usuario, el rol o el grupo puede asumir ese rol cuando se ejecute el comando. El permiso ASSUMEROLE permite conceder acceso a los comandos adecuados según sea necesario.

Solo un superusuario de base de datos puede conceder o revocar el permiso ASSUMEROLE para usuarios, roles y grupos. Un superusuario siempre retiene el permiso ASSUMEROLE.

Para habilitar el uso del permiso ASSUMEROLE por parte de usuarios, roles y grupos, un superusuario realiza las dos acciones siguientes:

- Ejecuta una vez la siguiente instrucción en el clúster:

```
revoke assumerole on all from public for all;
```

- Concede el permiso ASSUMEROLE a usuarios, roles y grupos para los comandos adecuados.

Puede especificar el encadenamiento de roles en la cláusula ON cuando concede el permiso ASSUMEROLE. Utiliza comas para separar los roles en una cadena de roles; por ejemplo, `Role1,Role2,Role3`. Si se especificó el encadenamiento de roles durante la concesión del permiso ASSUMEROLE, deberá especificar la cadena de roles cuando realice operaciones concedidas por el permiso ASSUMEROLE. No se pueden especificar roles individuales en la cadena de roles cuando se realizan operaciones concedidas por el permiso ASSUMEROLE. Por ejemplo, si a un usuario, un rol o un grupo se le concede la cadena de roles `Role1,Role2,Role3`, no se puede especificar solo `Role1` para realizar operaciones.

Si un usuario intenta realizar una operación COPY, UNLOAD, EXTERNAL FUNCTION o CREATE MODEL, y no se le ha concedido el permiso ASSUMEROLE, aparecerá un mensaje similar al siguiente.

```
ERROR:  User awsuser does not have ASSUMEROLE permission on IAM role
 "arn:aws:iam::123456789012:role/RoleA" for COPY
```

Para enumerar los usuarios a los que se les concedió acceso a roles de IAM y comandos a través del permiso ASSUMEROLE, consulte HAS_ASSUMEROLE_PRIVILEGE. Para enumerar los roles de IAM y los permisos de comando que se han concedido a un usuario que usted especifica, consulte PG_GET_IAM_ROLE_BY_USER. Para enumerar los usuarios, los roles y los grupos a los que se les concedió acceso a un rol de IAM que usted especificó, consulte PG_GET_GRANTEE_BY_IAM_ROLE.

Notas de uso para conceder permisos de machine learning

No puede conceder ni revocar directamente permisos relacionados con una función de ML. Una función de ML pertenece a un modelo de ML y los permisos se controlan mediante el modelo. En su lugar, puede conceder permisos relacionados con el modelo de ML. En el siguiente ejemplo, se muestra cómo conceder permisos a todos los usuarios para ejecutar la función de ML asociada al modelo `customer_churn`.

```
GRANT EXECUTE ON MODEL customer_churn TO PUBLIC;
```

También puede conceder todos los permisos a un usuario para el modelo de ML `customer_churn`.

```
GRANT ALL on MODEL customer_churn TO ml_user;
```

Se producirá un error en la concesión del permiso `EXECUTE` relacionado con una función de ML si existe una función de ML en el esquema, aunque dicha función de ML ya disponga del permiso `EXECUTE` mediante `GRANT EXECUTE ON MODEL`. Recomendamos utilizar un esquema independiente cuando utilice el comando `CREATE MODEL` para mantener las funciones de ML en un esquema independiente por sí mismas. En el siguiente ejemplo, se muestra cómo hacerlo.

```
CREATE MODEL ml_schema.customer_churn
FROM customer_data
TARGET churn
FUNCTION ml_schema.customer_churn_prediction
IAM_ROLE default
SETTINGS (
 S3_BUCKET 'amzn-s3-demo-bucket'
);
```

## Ejemplos

En el siguiente ejemplo, se le concede el privilegio SELECT en la tabla SALES al usuario `fred`.

```
grant select on table sales to fred;
```

En el siguiente ejemplo, se le concede el privilegio SELECT en todas las tablas del esquema QA_TICKIT al usuario `fred`.

```
grant select on all tables in schema qa_tickit to fred;
```

En el siguiente ejemplo, se le conceden todos los privilegios del esquema en el esquema QA_TICKIT al grupo de usuarios QA_USERS. Los privilegios de esquema son CREATE y USAGE. USAGE concede a los usuarios acceso a los objetos del esquema, pero no les concede privilegios (como INSERT o SELECT) sobre esos objetos. Otorga privilegios en cada objeto por separado.

```
create group qa_users;
grant all on schema qa_tickit to group qa_users;
```

En el siguiente ejemplo, se le conceden todos los privilegios de la tabla SALES en el esquema QA_TICKIT a todos los usuarios del grupo QA_USERS.

```
grant all on table qa_tickit.sales to group qa_users;
```

En el siguiente ejemplo, se conceden todos los privilegios en la tabla SALES del esquema QA_TICKIT a todos los usuarios de los grupos QA_USERS y RO_USERS.

```
grant all on table qa_tickit.sales to group qa_users, group ro_users;
```

En el siguiente ejemplo, se concede el privilegio DROP en la tabla SALES del esquema QA_TICKIT a todos los usuarios del grupo QA_USERS.

```
grant drop on table qa_tickit.sales to group qa_users;>
```

En la siguiente secuencia de comandos se muestra que obtener acceso a un esquema no concede privilegios en una tabla del esquema.

```
create user schema_user in group qa_users password 'Abcd1234';
create schema qa_tickit;
create table qa_tickit.test (col1 int);
```

```
grant all on schema qa_tickit to schema_user;

set session authorization schema_user;
select current_user;


current_user
--------------
schema_user
(1 row)


select count(*) from qa_tickit.test;


ERROR: permission denied for relation test [SQL State=42501]


set session authorization dw_user;
grant select on table qa_tickit.test to schema_user;
set session authorization schema_user;
select count(*) from qa_tickit.test;


count
-------
0
(1 row)
```

En la siguiente secuencia de comandos se muestra que obtener acceso a una vista no implica el acceso a sus tablas subyacentes. El usuario denominado VIEW_USER no puede seleccionar desde la tabla DATE aunque haya recibido todos los privilegios en VIEW_DATE.

```
create user view_user password 'Abcd1234';
create view view_date as select * from date;
grant all on view_date to view_user;
set session authorization view_user;
select current_user;


current_user
--------------
view_user
```

```
(1 row)


select count(*) from view_date;


count
-------
365
(1 row)


select count(*) from date;


ERROR:  permission denied for relation date
```

En el siguiente ejemplo, se concede el privilegio SELECT en las columnas `cust_name` y `cust_phone` de la tabla `cust_profile` al usuario `user1`.

```
grant select(cust_name, cust_phone) on cust_profile to user1;
```

En el siguiente ejemplo, se concede el privilegio SELECT en las columnas `cust_name` y `cust_phone` y el privilegio UPDATE en la columna `cust_contact_preference` de la tabla `cust_profile` al grupo `sales_group`.

```
grant select(cust_name, cust_phone), update(cust_contact_preference) on cust_profile to
 group sales_group;
```

En el siguiente ejemplo, se muestra el uso de la palabra clave ALL para otorgar privilegios SELECT y UPDATE en tres columnas de la tabla `cust_profile` al grupo `sales_admin`.

```
grant ALL(cust_name, cust_phone,cust_contact_preference) on cust_profile to group
 sales_admin;
```

En el siguiente ejemplo, se concede el privilegio SELECT en la columna `cust_name` de la vista `cust_profile_vw` al usuario `user2`.

```
grant select(cust_name) on cust_profile_vw to user2;
```

Ejemplos de concesión de acceso a recursos compartidos de datos

En los siguientes ejemplos, se muestran permisos de uso compartido de datos GRANT en una base de datos o un esquema específicos creados a partir de un datashare.

En el siguiente ejemplo, un administrador del lado del productor concede el permiso USAGE del recurso compartido de datos `salesshare` al espacio de nombres especificado.

```
GRANT USAGE ON DATASHARE salesshare TO NAMESPACE
 '13b8833d-17c6-4f16-8fe4-1a018f5ed00d';
```

En el siguiente ejemplo, un administrador del lado del consumidor concede el permiso USAGE en `sales_db` a `Bob`.

```
GRANT USAGE ON DATABASE sales_db TO Bob;
```

En el siguiente ejemplo, un administrador del lado del consumidor concede el permiso GRANT USAGE en el esquema `sales_schema` al rol `Analyst_role`. `sales_schema` es un esquema externo que apunta a sales_db.

```
GRANT USAGE ON SCHEMA sales_schema TO ROLE Analyst_role;
```

En este punto, `Bob` y `Analyst_role` pueden acceder a todos los objetos de la base de datos en `sales_schema` y `sales_db`.

El siguiente ejemplo muestra la concesión de permisos adicionales en el nivel de objeto para los objetos de una base de datos compartida. Estos permisos adicionales solo son necesarios si el comando CREATE DATABASE que se ha utilizado para crear la base de datos compartida ha utilizado la cláusula WITH PERMISSIONS. Si el comando CREATE DATABASE no ha utilizado WITH PERMISSIONS, al conceder USAGE en la base de datos compartida se otorga acceso total a todos los objetos de esa base de datos.

```
GRANT SELECT ON sales_db.sales_schema.tickit_sales_redshift to Bob;
```

Ejemplos de concesión de permisos acotados

El siguiente ejemplo concede el uso de todos los esquemas actuales y futuros de la base de datos `Sales_db` al rol `Sales`.

```
GRANT USAGE FOR SCHEMAS IN DATABASE Sales_db TO ROLE Sales;
```

El siguiente ejemplo otorga el permiso SELECT para todas las tablas actuales y futuras de la base de datos `Sales_db` al usuario `alice` y también otorga a `alice` el permiso para conceder permisos acotados en las tablas `Sales_db` a otros usuarios.

```
GRANT SELECT FOR TABLES IN DATABASE Sales_db TO alice WITH GRANT OPTION;
```

En el siguiente ejemplo, se concede el permiso EXECUTE para las funciones del esquema `Sales_schema` al usuario bob.

```
GRANT EXECUTE FOR FUNCTIONS IN SCHEMA Sales_schema TO bob;
```

En el siguiente ejemplo se conceden todos los permisos para todas las tablas del esquema `ShareSchema` de la base de datos `ShareDb` al rol `Sales`. Al especificar el esquema, puede indicar la base de datos del esquema mediante el formato de dos partes `database.schema`.

```
GRANT ALL FOR TABLES IN SCHEMA ShareDb.ShareSchema TO ROLE Sales;
```

El siguiente ejemplo es el mismo que el anterior. Puede especificar la base de datos mediante la palabra clave `DATABASE` en lugar de utilizar un formato de dos partes.

```
GRANT ALL FOR TABLES IN SCHEMA ShareSchema DATABASE ShareDb TO ROLE Sales;
```

Ejemplos de concesión del privilegio ASSUMEROLE

A continuación, se muestran ejemplos de concesión del privilegio ASSUMEROLE.

En el siguiente ejemplo, se muestra la instrucción REVOKE que un superusuario ejecuta una vez en el clúster con objeto de habilitar el uso del privilegio ASSUMEROLE para usuarios y grupos. Luego, el superusuario concede el privilegio ASSUMEROLE a los usuarios y los grupos para los comandos adecuados. Para obtener información sobre cómo habilitar el uso del privilegio ASSUMEROLE para usuarios y grupos, consulte Notas de uso para conceder el permiso ASSUMEROLE.

```
revoke assumerole on all from public for all;
```

En el siguiente ejemplo, se concede el privilegio ASSUMEROLE al usuario `reg_user1` para el rol de IAM `Redshift-S3-Read` con objeto de realizar operaciones COPY.

```
grant assumerole on 'arn:aws:iam::123456789012:role/Redshift-S3-Read'
to reg_user1 for copy;
```

En el siguiente ejemplo, se concede el privilegio ASSUMEROLE al usuario `reg_user1` para la cadena de roles de IAM `RoleA`, `RoleB` con objeto de realizar operaciones UNLOAD.

```
grant assumerole
on 'arn:aws:iam::123456789012:role/RoleA,arn:aws:iam::210987654321:role/RoleB'
to reg_user1
for unload;
```

A continuación, se muestra un ejemplo del comando UNLOAD a través de la cadena de roles de IAM `RoleA`, `RoleB`.

```
unload ('select * from venue limit 10')
to 's3://companyb/redshift/venue_pipe_'
iam_role 'arn:aws:iam::123456789012:role/RoleA,arn:aws:iam::210987654321:role/RoleB';
```

En el siguiente ejemplo, se concede el privilegio ASSUMEROLE al usuario `reg_user1` para el rol de IAM `Redshift-Exfunc` con objeto de crear funciones externas.

```
grant assumerole on 'arn:aws:iam::123456789012:role/Redshift-Exfunc'
to reg_user1 for external function;
```

En el siguiente ejemplo, se concede el privilegio ASSUMEROLE al usuario `reg_user1` para el rol de IAM `Redshift-model` con objeto de crear modelos de machine learning.

```
grant assumerole on 'arn:aws:iam::123456789012:role/Redshift-ML'
to reg_user1 for create model;
```

Ejemplos de concesión de los privilegios ROLE

El siguiente ejemplo concede sample_role1 a user1.

```
CREATE ROLE sample_role1;
GRANT ROLE sample_role1 TO user1;
```

El siguiente ejemplo concede sample_role1 a user1 con WITH ADMIN OPTION, establece la sesión actual para user1, y user1 concede sample_role1 a user2.

```
GRANT ROLE sample_role1 TO user1 WITH ADMIN OPTION;
SET SESSION AUTHORIZATION user1;
GRANT ROLE sample_role1 TO user2;
```

El siguiente ejemplo concede sample_role1 a sample_role2.

```
GRANT ROLE sample_role1 TO ROLE sample_role2;
```

El siguiente ejemplo concede sample_role2 a sample_role3 y sample_role4. Luego, intenta conceder sample_role3 a sample_role1.

```
GRANT ROLE sample_role2 TO ROLE sample_role3;
GRANT ROLE sample_role3 TO ROLE sample_role2;
ERROR: cannot grant this role, a circular dependency was detected between these roles
```

El siguiente ejemplo concede los privilegios del sistema CREATE USER a sample_role1.

```
GRANT CREATE USER TO ROLE sample_role1;
```

El siguiente ejemplo concede el rol definido por el sistema `sys:dba` a user1.

```
GRANT ROLE sys:dba TO user1;
```

El siguiente ejemplo intenta conceder sample_role3 en una dependencia circular a sample_role2.

```
CREATE ROLE sample_role3;
GRANT ROLE sample_role2 TO ROLE sample_role3;
GRANT ROLE sample_role3 TO ROLE sample_role2; -- fail
ERROR:  cannot grant this role, a circular dependency was detected between these roles
```

# INSERT

Temas


- Sintaxis
- Parámetros
- Notas de uso
- Ejemplos de INSERT

Inserta nuevas filas en una tabla. Puede insertar una única fila con la sintaxis VALUES, varias filas con la sintaxis VALUES, o una o más filas definidas por los resultados de una consulta (INSERT INTO...SELECT).

> **Note**
>
> Le recomendamos usar el comando COPY para cargar grandes cantidades de datos. El uso de instrucciones INSERT individuales para completar una tabla podría ser prohibitivamente lento. Además, si los datos ya existen en otras tablas de la base de datos de Amazon Redshift, utilice INSERT INTO SELECT o CREATE TABLE AS para mejorar el rendimiento. Para obtener más información acerca del comando COPY para cargar tablas, consulte Carga de datos en Amazon Redshift.

> **Note**
>
> El tamaño máximo de una instrucción SQL es de 16 MB.

## Sintaxis

```
INSERT INTO table_name [ ( column [, ...] ) ]
{DEFAULT VALUES |
VALUES ( { expression | DEFAULT } [, ...] )
[, ( { expression | DEFAULT } [, ...] )
[, ...] ] |
query }
```

## Parámetros

table_name

Una tabla temporal o persistente. Solo el propietario de la tabla o un usuario con el privilegio INSERT en la tabla puede insertar filas. Si utiliza la cláusula query para insertar filas, debe tener el privilegio SELECT en las tablas que aparecen en la consulta.

> **Note**
>
> Utilice INSERT (tabla externa) para insertar los resultados de una consulta SELECT en tablas existentes del catálogo externo. Para obtener más información, consulte INSERT (tabla externa).

column

Puede insertar valores en una o más columnas de la tabla. Puede enumerar los nombres de las columnas de destino en cualquier orden. Si no se especifica una lista de columnas, los valores que se inserten deben coincidir con las columnas de la tabla en el orden en que se declararon en la instrucción CREATE TABLE. Si la cantidad de valores que deben insertarse es menor que la cantidad de columnas de la tabla, se cargarán las primeras n columnas.

O bien se carga el valor predeterminado declarado, o bien se carga un valor nulo en cualquier columna que no aparezca (de forma implícita o explícita) en la instrucción INSERT.

DEFAULT VALUES

Si se asignaron valores predeterminados a las columnas de la tabla cuando se creó la tabla, utilice estas palabras clave para insertar una fila que conste completamente de valores predeterminados. Si ninguna de las columnas tiene valores predeterminados, se insertan valores nulos en esas columnas. Si alguna de las columnas se declara NOT NULL, la instrucción INSERT devuelve un error.

VALUES

Utilice esta palabra clave para insertar una o más filas; cada fila debe constar de uno o más valores. La lista VALUES para cada fila debe alinearse con la lista de columnas. Para insertar varias filas, utilice un delimitador de comas entre cada lista de expresiones. No repita la palabra clave VALUES. Todas las listas VALUES para una instrucción INSERT de varias filas deben contener la misma cantidad de valores.

expresión

Un valor único o una expresión que toma el valor de un valor único. Cada valor debe ser compatible con el tipo de datos de la columna donde se inserta. Si es posible, los valores que tengan un tipo de datos que no coincida con el tipo de datos declarado en la columna se convertirán automáticamente en un tipo de datos compatible. Por ejemplo:

- Un valor decimal `1.1` se inserta en una columna INT como 1.

- Un valor decimal `100.8976` se inserta en una columna DEC (5,2) como `100.90`.

Puede convertir explícitamente un valor a un tipo de datos compatible al incluir la sintaxis de conversión de tipo en la expresión. Por ejemplo, si la columna COL1 de la tabla T1 es una columna CHAR(3):

```
insert into t1(col1) values('Incomplete'::char(3));
```

Esta instrucción inserta el valor `Inc` en la columna.

Para una instrucción INSERT VALUES de una fila, puede usar una subconsulta escalar como expresión. El resultado de la subconsulta se inserta en la columna adecuada.

> **Note**
>
> Las subconsultas no pueden utilizarse como expresiones de instrucciones INSERT VALUES de varias filas.

DEFAULT

Utilice esta palabra clave para insertar el valor predeterminado de una columna, como se definió cuando se creó la tabla. Si no existe un valor predeterminado para una columna, se inserta un valor nulo. No se puede insertar un valor predeterminado en una columna que tenga una restricción NOT NULL si esa columna no tiene un valor predeterminado explícito asignado en la instrucción CREATE TABLE.

consulta

Inserte una o más filas en la tabla al definir cualquier consulta. Todas las filas que la consulta produce se insertan en la tabla. La consulta debe devolver una lista de columnas compatible con las columnas de la tabla, pero los nombres de las columnas no tienen que coincidir.

## Notas de uso

> **Note**
>
> Le recomendamos usar el comando COPY para cargar grandes cantidades de datos. El uso de instrucciones INSERT individuales para completar una tabla podría ser

prohibitivamente lento. Además, si los datos ya existen en otras tablas de la base de datos de Amazon Redshift, utilice INSERT INTO SELECT o CREATE TABLE AS para mejorar el rendimiento. Para obtener más información acerca del comando COPY para cargar tablas, consulte Carga de datos en Amazon Redshift.

El formato de datos para los valores insertados debe coincidir con el formato de datos especificado por la definición CREATE TABLE.

Después de insertar una gran cantidad de filas nuevas en una tabla:

- Limpie la tabla para recuperar espacio de almacenamiento y reordene las filas.
- Analice la tabla para actualizar las estadísticas para el planificador de consultas.

Cuando los valores se insertan en columnas DECIMAL y superan la escala especificada, los valores cargados se redondean según sea adecuado. Por ejemplo, cuando un valor de `20.259` se inserta en una columna DECIMAL(8,2), el valor que se almacena es `20.26`.

Puede insertar datos en una columna GENERATED BY DEFAULT AS IDENTITY. Puede actualizar las columnas definidas como GENERATED BY DEFAULT AS IDENTITY con los valores proporcionados. Para obtener más información, consulte GENERATED BY DEFAULT AS IDENTITY.

## Ejemplos de INSERT

La tabla CATEGORY de la base de datos TICKIT contiene las siguientes filas:

```
 catid | catgroup |  catname  |                 catdesc
-------+----------+-----------+--------------------------------------------
     1 | Sports   | MLB       | Major League Baseball
     2 | Sports   | NHL       | National Hockey League
     3 | Sports   | NFL       | National Football League
     4 | Sports   | NBA       | National Basketball Association
     5 | Sports   | MLS       | Major League Soccer
     6 | Shows    | Musicals  | Musical theatre
     7 | Shows    | Plays     | All non-musical theatre
     8 | Shows    | Opera     | All opera and light opera
     9 | Concerts | Pop       | All rock and pop music concerts
    10 | Concerts | Jazz      | All jazz singers and bands
    11 | Concerts | Classical | All symphony, concerto, and choir concerts
(11 rows)
```

Cree una tabla CATEGORY_STAGE con un esquema similar a la tabla CATEGORY, pero defina valores predeterminados para las columnas:

```
create table category_stage
(catid smallint default 0,
catgroup varchar(10) default 'General',
catname varchar(10) default 'General',
catdesc varchar(50) default 'General');
```

La siguiente instrucción INSERT selecciona todas las filas de la tabla CATEGORY y las inserta en la tabla CATEGORY_STAGE.

```
insert into category_stage
(select * from category);
```

Los paréntesis alrededor de la consulta son opcionales.

Este comando inserta una nueva fila en la tabla CATEGORY_STAGE con un valor especificado para cada columna en orden:

```
insert into category_stage values
(12, 'Concerts', 'Comedy', 'All stand-up comedy performances');
```

También puede insertar una nueva fila que combine valores específicos y valores predeterminados:

```
insert into category_stage values
(13, 'Concerts', 'Other', default);
```

Ejecute la siguiente consulta para devolver las filas insertadas:

```
select * from category_stage
where catid in(12,13) order by 1;

 catid | catgroup | catname |             catdesc
-------+----------+---------+---------------------------------
    12 | Concerts | Comedy  | All stand-up comedy performances
    13 | Concerts | Other   | General
(2 rows)
```

En los siguientes ejemplos se muestran algunas instrucciones INSERT VALUES de varias filas. El primer ejemplo inserta valores CATID específicos para dos filas y los valores predeterminados para las otras columnas de ambas filas.

```
insert into category_stage values
(14, default, default, default),
(15, default, default, default);

select * from category_stage where catid in(14,15) order by 1;
 catid | catgroup | catname | catdesc
-------+----------+---------+---------
    14 | General  | General | General
    15 | General  | General | General
(2 rows)
```

El próximo ejemplo inserta tres filas con diversas combinaciones de valores predeterminados y específicos:

```
insert into category_stage values
(default, default, default, default),
(20, default, 'Country', default),
(21, 'Concerts', 'Rock', default);

select * from category_stage where catid in(0,20,21) order by 1;
 catid | catgroup | catname | catdesc
-------+----------+---------+---------
     0 | General  | General | General
    20 | General  | Country | General
    21 | Concerts | Rock    | General
(3 rows)
```

El primer conjunto de VALUES en este ejemplo produce los mismos resultados que especificar DEFAULT VALUES para una instrucción INSERT de una sola fila.

En los siguientes ejemplos se muestra un comportamiento INSERT cuando una tabla tiene una columna IDENTITY. Primero, cree una nueva versión de la tabla CATEGORY, luego inserte filas en la tabla desde CATEGORY:

```
create table category_ident
(catid int identity not null,
catgroup varchar(10) default 'General',
```

```
catname varchar(10) default 'General',
catdesc varchar(50) default 'General');


insert into category_ident(catgroup,catname,catdesc)
select catgroup,catname,catdesc from category;
```

Tenga en cuenta que no puede insertar valores enteros específicos en la columna CATID IDENTITY. Los valores de la columna IDENTITY se generan automáticamente.

En el siguiente ejemplo, se demuestra que las subconsultas no pueden utilizarse como expresiones en instrucciones INSERT VALUES de varias filas:

```
insert into category(catid) values
((select max(catid)+1 from category)),
((select max(catid)+2 from category));

ERROR: can't use subqueries in multi-row VALUES
```

En el siguiente ejemplo se muestra una inserción en una tabla temporal rellena con datos de la tabla `venue` mediante la cláusula `WITH SELECT`. Para obtener más información sobre la tabla `venue`, consulte Base de datos de muestra.

Primero, cree la tabla temporal `#venuetemp`.

```
CREATE TABLE #venuetemp AS SELECT * FROM venue;
```

Muestre las filas de la tabla `#venuetemp`.

```
SELECT * FROM #venuetemp ORDER BY venueid;

venueid | venuename                | venuecity  | venuestate| venueseats
--------+--------------------------+------------+-----------+------------
1       Toyota Park                Bridgeview   IL          0
2       Columbus Crew Stadium      Columbus     OH          0
3       RFK Stadium                Washington   DC          0
4       CommunityAmerica Ballpark  Kansas City  KS          0
5       Gillette Stadium           Foxborough   MA          68756
...
```

Inserte 10 filas duplicadas en la tabla `#venuetemp` mediante la cláusula `WITH SELECT`.

```
INSERT INTO #venuetemp (WITH venuecopy AS (SELECT * FROM venue) SELECT * FROM venuecopy
 ORDER BY 1 LIMIT 10);
```

Muestre las filas de la tabla #venuetemp.

```
SELECT * FROM #venuetemp ORDER BY venueid;

venueid | venuename                | venuecity  | venuestate| venueseats
--------+--------------------------+------------+-----------+------------
1       Toyota Park                Bridgeview   IL          0
1       Toyota Park                Bridgeview   IL          0
2       Columbus Crew Stadium      Columbus     OH          0
2       Columbus Crew Stadium      Columbus     OH          0
3       RFK Stadium                Washington   DC          0
3       RFK Stadium                Washington   DC          0
4       CommunityAmerica Ballpark  Kansas City  KS          0
4       CommunityAmerica Ballpark  Kansas City  KS          0
5       Gillette Stadium           Foxborough   MA          68756
5       Gillette Stadium           Foxborough   MA          68756
...
```

## INSERT (tabla externa)

Inserta los resultados de una consulta SELECT en tablas externas existentes en un catálogo externo, como para AWS Glue, AWS Lake Formation o un metastore de Apache Hive. Utilice el mismo rol de AWS Identity and Access Management (IAM) utilizado para el comando CREATE EXTERNAL SCHEMA para interactuar con catálogos externos y Amazon S3.

Para las tablas sin partición, el comando INSERT (tabla externa) escribe los datos en la ubicación de Amazon S3 definida en la tabla, en función de las propiedades de la tabla y el formato de archivo especificados.

Para las tablas con partición, INSERT (tabla externa) escribe datos en la ubicación de Amazon S3 de acuerdo con la clave de partición especificada en la tabla. También registra automáticamente nuevas particiones en el catálogo externo una vez finalizada la operación INSERT.

No se puede ejecutar INSERT (tabla externa) en un bloque de transacción (BEGIN … END). Para obtener más información acerca de las transacciones, consulte Aislamiento serializable.

## Sintaxis

```
INSERT INTO external_schema.table_name
{ select_statement }
```

## Parámetros

external_schema.table_name

El nombre de un esquema externo existente y una tabla externa de destino en la que se va a insertar.

select_statement

Una instrucción que inserta una o más filas en la tabla externa definiendo cualquier consulta. Todas las filas que produce la consulta se escriben en Amazon S3 en formato de texto o Parquet según la definición de la tabla. La consulta debe devolver una lista de columnas que sea compatible con los tipos de datos de columna de la tabla externa. No obstante, los nombres de columna no tienen que coincidir.

## Notas de uso

El número de columnas de la consulta SELECT debe ser el mismo que la suma de las columnas de datos y las columnas de partición. La ubicación y el tipo de datos de cada columna de datos deben coincidir con los de la tabla externa. La ubicación de las columnas de partición debe estar al final de la consulta SELECT, en el mismo orden en que se definieron en el comando CREATE EXTERNAL TABLE. Los nombres de columna no tienen que coincidir.

En algunos casos, es posible que desee ejecutar el comando INSERT (tabla externa) en un catálogo de datos de AWS Glue o en un metastore de Hive. En el caso de AWS Glue, el rol de IAM utilizado para crear el esquema externo debe tener permisos de lectura y escritura en Amazon S3 y AWS Glue. Si utiliza un catálogo de AWS Lake Formation, este rol de IAM se convierte en el propietario de la nueva tabla de Lake Formation. Este rol de IAM debe tener al menos los permisos siguientes:

- Permisos SELECT, INSERT, UPDATE en la tabla externa
- Permiso de ubicación de datos en la ruta de Amazon S3 de la tabla externa

Para asegurarse de que los nombres de los archivos son únicos, Amazon Redshift utiliza el siguiente formato para el nombre de cada archivo cargado en Amazon S3 de manera predeterminada.

*<date>*_*<time>*_*<microseconds>*_*<query_id>*_*<slice-number>*_part_*<part-number>*.*<format>*.

Un ejemplo es 20200303_004509_810669_1007_0001_part_00.parquet.

Tenga en cuenta lo siguiente al ejecutar el comando INSERT (tabla externa):

- No se admiten tablas externas que tengan un formato distinto de PARQUET o TEXTFILE.
- Este comando admite propiedades de tabla existentes como 'write.parallel', 'write.maxfilesize.mb', 'compression_type' y 'serialization.null.format'. Para actualizar esos valores, ejecute el comando ALTER TABLE SET TABLE PROPERTIES.
- La propiedad de tabla 'numRows' se actualiza automáticamente hacia el final de la operación INSERT. La propiedad de tabla debe definirse o agregarse a la tabla si no se creó mediante la operación CREATE EXTERNAL TABLE AS.
- La cláusula LIMIT no se admite en la consulta SELECT externa. En su lugar, utilice una cláusula LIMIT anidada.
- Puede utilizar la tabla STL_UNLOAD_LOG para realizar un seguimiento de los archivos que se escribieron en Amazon S3 mediante cada operación INSERT (tabla externa).
- Amazon Redshift solo es compatible con el cifrado estándar de Amazon S3 para INSERT (tabla externa).

## Ejemplos de INSERT (tabla externa)

En el siguiente ejemplo, se insertan los resultados de la instrucción SELECT en la tabla externa.

```
INSERT INTO spectrum.lineitem
SELECT * FROM local_lineitem;
```

En el siguiente ejemplo, se insertan los resultados de la instrucción SELECT en una tabla externa particionada mediante la partición estática. Las columnas de partición están codificadas de forma rígida en la instrucción SELECT. Las columnas de partición deben estar al final de la consulta.

```
INSERT INTO spectrum.customer
SELECT name, age, gender, 'May', 28 FROM local_customer;
```

En el siguiente ejemplo, se insertan los resultados de la instrucción SELECT en una tabla externa particionada mediante la partición dinámica. Las columnas de partición no están codificadas de forma

rígida. Los datos se agregan automáticamente a las carpetas de partición existentes o a las nuevas carpetas si se agrega una nueva partición.

```
INSERT INTO spectrum.customer
SELECT name, age, gender, month, day FROM local_customer;
```

# LOCK

Restringe el acceso a una tabla de la base de datos. Este comando solo es significativo cuando se ejecuta dentro de un bloque de transacciones.

El comando LOCK obtiene un bloqueo del nivel de la tabla en el modo "ACCESS EXCLUSIVE", en el que se espera, si es necesario, que se libere cualquier bloqueo en conflicto. Bloquear una tabla explícitamente en esta manera hace que las lecturas y escrituras de la tabla esperen cuando se las intenta desde otras transacciones o sesiones. Un bloqueo de tabla explícito creado por un usuario temporalmente previene que otro usuario seleccione datos de la tabla o cargue datos en ella. El bloqueo se libera cuando se completa la transacción que contiene el comando LOCK.

Los bloqueos de tabla menos restrictivos son adquiridos implícitamente por comandos que hacen referencia a tablas, como operaciones de escritura. Por ejemplo, si un usuario intenta leer datos de una tabla mientras otro usuario actualiza la tabla, los datos que se leen serán una instantánea de los datos que ya se confirmaron. (En algunos casos, las consultas se detienen si infringen las reglas de aislamiento serializable). Consulte Administración de operaciones de escritura simultáneas.

Algunas operaciones DDL, como DROP TABLE y TRUNCATE, crean bloqueos exclusivos. Estas operaciones previenen las lecturas de datos.

Si se produce un conflicto de bloqueo, Amazon Redshift muestra un mensaje de error para alertar al usuario que inició la transacción en conflicto. La transacción que recibió el conflicto de bloqueo se detiene. Cada vez que se produce un conflicto de bloqueo, Amazon Redshift escribe una entrada en la tabla STL_TR_CONFLICT.

## Sintaxis

```
LOCK [ TABLE ] table_name [, ...]
```

### Parámetros

TABLE

Palabra clave opcional.

table_name

Nombre de la tabla que se bloqueará. Es posible bloquear más de una tabla al utilizar una lista de nombres de tabla delimitada por comas. No es posible bloquear vistas.

### Ejemplo

```
begin;

lock event, sales;

...
```

## MERGE

Fusiona de forma condicional las filas de una tabla de origen con una tabla de destino. Tradicionalmente, esto solo puede conseguirse utilizando varias sentencias de inserción, actualización o eliminación por separado. Para obtener más información sobre las operaciones que MERGE le permite combinar, consulte UPDATE, DELETE e INSERT.

### Sintaxis

```
MERGE INTO target_table
USING source_table [ [ AS ] alias ]
ON match_condition
[ WHEN MATCHED THEN { UPDATE SET col_name = { expr } [,...] | DELETE }
WHEN NOT MATCHED THEN INSERT [ ( col_name [,...] ) ] VALUES ( { expr } [, ...] ) |
REMOVE DUPLICATES ]
```

### Parámetros

target_table

La tabla temporal o permanente en la que se fusiona la instrucción MERGE.

source_table

Tabla temporal o permanente que proporciona las filas que se van a fusionar en target_table. source_table también puede ser una tabla de Spectrum.

alias

El nombre alternativo temporal de source_table.

Este parámetro es opcional. El alias anterior con AS también es opcional.

match_condition

Especifica los predicados iguales entre la columna de la tabla de origen y la columna de la tabla de destino que se utilizan para determinar si las filas de source_table pueden coincidir con las filas de target_table. Si se cumple la condición, MERGE ejecuta matched_clause para esa fila. De lo contrario, MERGE ejecuta not_matched_clause para esa fila.

WHEN MATCHED

Especifica la acción que se ejecutará cuando la condición de coincidencia entre una fila de origen y una fila de destino se evalúe como True. Puede especificar una acción UPDATE o una acción DELETE.

UPDATE

Actualiza la fila coincidente en target_table. Solo se actualizan los valores de col_name que especifique.

DELETE

Elimina la fila coincidente en target_table.

WHEN NOT MATCHED

Especifica la acción que se ejecutará cuando la condición de coincidencia se evalúe como False o Unknown. Solo puede especificar la acción de inserción INSERT para esta cláusula.

INSERT

Inserta en target_table filas de source_table que no coinciden con ninguna fila de target_table, según la condición de match_condition. El valor de col_name de destino puede enumerarse en cualquier orden. Si no proporciona ningún valor de col_name, el orden predeterminado será el de todas las columnas de la tabla en su orden declarado.

col_name

Una o más columnas que desea modificar. No incluya el nombre de la tabla al especificar la columna de destino.

expr

La expresión que define el nuevo valor de col_name.

REMOVE DUPLICATES

Especifica que el comando MERGE se ejecuta en modo simplificado. El modo simplificado tiene los siguientes requisitos:

- target_table y source_table deben tener el mismo número de columnas y tipos de columna compatibles.
- Omita la cláusula WHEN y las cláusulas UPDATE e INSERT de su comando MERGE.
- Utilice la cláusula REMOVE DUPLICATES en su comando MERGE.

En el modo simplificado, MERGE hace lo siguiente:

- Las filas de target_table que tienen una coincidencia en source_table se actualizan para que coincidan con los valores de source_table.
- Las filas de source_table que no tengan una coincidencia en target_table se insertan en target_table.
- Cuando varias filas de target_table coinciden con la misma fila de source_table, se eliminan las filas duplicadas. Amazon Redshift mantiene una fila y la actualiza. Las filas duplicadas que no coinciden con una fila de source_table permanecen sin cambios.

El uso de REMOVE DUPLICATES ofrece mejor rendimiento que usar WHEN MATCHED y WHEN NOT MATCHED. Recomendamos utilizar REMOVE DUPLICATES si target_table y source_table son compatibles y no necesita conservar filas duplicadas en target_table.

## Notas de uso

- Para ejecutar instrucciones MERGE, debe ser el propietario tanto de source_table como de target_table o tener el permiso SELECT para esas tablas. Además, debe tener los permisos UPDATE, DELETE e INSERT para target_table, según las operaciones incluidas en la instrucción MERGE.
- target_table no puede ser una tabla del sistema, una tabla de catálogo ni una tabla externa.
- source_table y target_table no pueden ser la misma tabla.

- No se puede utilizar la cláusula WITH en una instrucción MERGE.
- Las filas de target_table no pueden coincidir con varias filas de source_table.

  Considere el siguiente ejemplo:

```
CREATE TABLE target (id INT, name CHAR(10));
CREATE TABLE source (id INT, name CHAR(10));

INSERT INTO target VALUES (1, 'Bob'), (2, 'John');
INSERT INTO source VALUES (1, 'Tony'), (1, 'Alice'), (3, 'Bill');

MERGE INTO target USING source ON target.id = source.id
WHEN MATCHED THEN UPDATE SET id = source.id, name = source.name
WHEN NOT MATCHED THEN INSERT VALUES (source.id, source.name);
ERROR: Found multiple matches to update the same tuple.

MERGE INTO target USING source ON target.id = source.id
WHEN MATCHED THEN DELETE
WHEN NOT MATCHED THEN INSERT VALUES (source.id, source.name);
ERROR: Found multiple matches to update the same tuple.
```

  En ambas instrucciones MERGE, se produce un error en la operación porque hay varias filas en la tabla `source` con un valor de ID de 1.
- match_condition y expr no pueden hacer referencia de forma parcial a columnas de tipo SUPER. Por ejemplo, si el objeto de tipo SUPER es una matriz o una estructura, no puede usar elementos individuales de esa columna para match_condition ni expr, pero puede usar toda la columna.

  Considere el siguiente ejemplo:

```
CREATE TABLE IF NOT EXISTS target (key INT, value SUPER);
CREATE TABLE IF NOT EXISTS source (key INT, value SUPER);

INSERT INTO target VALUES (1, JSON_PARSE('{"key": 88}'));
INSERT INTO source VALUES (1, ARRAY(1, 'John')), (2, ARRAY(2, 'Bill'));

MERGE INTO target USING source ON target.key = source.key
WHEN matched THEN UPDATE SET value = source.value[0]
WHEN NOT matched THEN INSERT VALUES (source.key, source.value[0]);
ERROR: Partial reference of SUPER column is not supported in MERGE statement.
```

  Para obtener más información sobre el tipo SUPER, consulte Tipo SUPER.

- Si source_table es grande, definir las columnas de combinación de target_table y source_table como claves de distribución puede mejorar el rendimiento.
- Para utilizar la cláusula REMOVE DUPLICATES, necesita permisos SELECT, INSERT y DELETE para target_table.
- source_table puede ser una vista o una subconsulta. A continuación, se muestra un ejemplo de una instrucción MERGE en la que source_table es una subconsulta que elimina las filas duplicadas.

```
MERGE INTO target
USING (SELECT id, name FROM source GROUP BY 1, 2) as my_source
ON target.id = my_source.id
WHEN MATCHED THEN UPDATE SET id = my_source.id, name = my_source.name
WHEN NOT MATCHED THEN INSERT VALUES (my_source.id, my_source.name);
```

- El destino no puede ser un origen de datos de ninguna subconsulta de la misma instrucción MERGE. Por ejemplo, el siguiente comando SQL devuelve un error como ERROR: Source view/ subquery in Merge statement cannot reference target table. porque la subconsulta hace referencia a `target` en vez de a `source`.

```
MERGE INTO target
USING (SELECT id, name FROM target GROUP BY 1, 2) as my_source
ON target.id = my_source.id
WHEN MATCHED THEN UPDATE SET id = my_source.id, name = my_source.name
WHEN NOT MATCHED THEN INSERT VALUES (my_source.id, my_source.name);
```

## Ejemplos

En el siguiente ejemplo, se crean dos tablas y, a continuación, se ejecuta una operación MERGE en ellas, se actualizan las filas coincidentes de la tabla de destino y se insertan las filas que no coinciden. A continuación, inserta otro valor en la tabla de origen y ejecuta otra operación MERGE, esta vez se eliminan las filas coincidentes y se inserta la nueva fila de la tabla de origen.

Primero cree y rellene las tablas de origen y destino.

```
CREATE TABLE target (id INT, name CHAR(10));
CREATE TABLE source (id INT, name CHAR(10));

INSERT INTO target VALUES (101, 'Bob'), (102, 'John'), (103, 'Susan');
INSERT INTO source VALUES (102, 'Tony'), (103, 'Alice'), (104, 'Bill');
```

```
SELECT * FROM target;
 id  |    name
-----+------------
 101 | Bob
 102 | John
 103 | Susan
(3 rows)

SELECT * FROM source;
 id  |    name
-----+------------
 102 | Tony
 103 | Alice
 104 | Bill
(3 rows)
```

A continuación, fusione la tabla de origen con la de destino, actualice la tabla de destino con las filas coincidentes e inserte las filas de la tabla de origen que no coincidan.

```
MERGE INTO target USING source ON target.id = source.id
WHEN MATCHED THEN UPDATE SET id = source.id, name = source.name
WHEN NOT MATCHED THEN INSERT VALUES (source.id, source.name);

SELECT * FROM target;
 id  |    name
-----+------------
 101 | Bob
 102 | Tony
 103 | Alice
 104 | Bill
(4 rows)
```

Tenga en cuenta que las filas con valores de identificación (id) de 102 y 103 se actualizan para que coincidan con los valores de nombre (name) de la tabla de destino. Además, se inserta una nueva fila con un valor de identificación de 104 y el valor de nombre de Bill en la tabla de destino.

A continuación, inserte una nueva fila en la tabla de origen.

```
INSERT INTO source VALUES (105, 'David');

SELECT * FROM source;
```

```
 id  |    name
-----+------------
 102 | Tony
 103 | Alice
 104 | Bill
 105 | David
(4 rows)
```

Por último, ejecute una operación de fusión para eliminar las filas coincidentes de la tabla de destino e insertar las filas que no coincidan.

```
MERGE INTO target USING source ON target.id = source.id
WHEN MATCHED THEN DELETE
WHEN NOT MATCHED THEN INSERT VALUES (source.id, source.name);

SELECT * FROM target;
 id  |    name
-----+------------
 101 | Bob
 105 | David
(2 rows)
```

Las filas con los valores de identificación 102, 103 y 104 se eliminan de la tabla de destino y se inserta una nueva fila con un valor de identificación de 105 y un valor de nombre de David en la tabla de destino.

En el siguiente ejemplo, se muestra la sintaxis simplificada de un comando MERGE que utiliza la cláusula REMOVE DUPLICATES.

```
CREATE TABLE target (id INT, name CHAR(10));
CREATE TABLE source (id INT, name CHAR(10));

INSERT INTO target VALUES (30, 'Tony'), (11, 'Alice'), (23, 'Bill');
INSERT INTO source VALUES (23, 'David'), (22, 'Clarence');

MERGE INTO target USING source ON target.id = source.id REMOVE DUPLICATES;

SELECT * FROM target;
id | name
---+------------
30 | Tony
11 | Alice
```

```
23 | David
22 | Clarence
(4 rows)
```

En el siguiente ejemplo, se muestra la sintaxis simplificada de un comando MERGE que utiliza la cláusula REMOVE DUPLICATES, lo que elimina las filas duplicadas de target_table si tienen filas coincidentes en source_table.

```
CREATE TABLE target (id INT, name CHAR(10));
CREATE TABLE source (id INT, name CHAR(10));

INSERT INTO target VALUES (30, 'Tony'), (30, 'Daisy'), (11, 'Alice'), (23, 'Bill'),
 (23, 'Nikki');
INSERT INTO source VALUES (23, 'David'), (22, 'Clarence');

MERGE INTO target USING source ON target.id = source.id REMOVE DUPLICATES;

SELECT * FROM target;
id | name
---+------------
30 | Tony
30 | Daisy
11 | Alice
23 | David
22 | Clarence
(5 rows)
```

Después de ejecutar MERGE, solo hay una fila con un valor de ID de 23 en target_table. Como no había ninguna fila en source_table con el valor de ID 30, las dos filas duplicadas con valores de ID 30 permanecen en target_table.

### Véase también

INSERT, UPDATE, DELETE

## PREPARE

Prepare una instrucción para su ejecución.

PREPARE crea una instrucción preparada. Cuando se ejecuta la instrucción PREPARE, la instrucción especificada (SELECT, INSERT, UPDATE o DELETE) se analiza, reescribe y planifica.

Cuando se emite un comando EXECUTE para la instrucción preparada, Amazon Redshift puede revisar el plan de ejecución de la consulta (para mejorar el rendimiento en función de los valores del parámetro especificado) antes de ejecutar la instrucción preparada.

## Sintaxis

```
PREPARE plan_name [ (datatype [, ...] ) ] AS statement
```

## Parámetros

plan_name

Un nombre arbitrario proporcionado a esta instrucción preparada particular. Debe ser único en el marco de una sesión y se utiliza posteriormente para ejecutar o anular la asignación de una instrucción preparada con anterioridad.

datatype (tipo_de_datos)

El tipo de datos de un parámetro para la instrucción preparada. Para hacer referencia a los parámetros de la instrucción preparada, use 1 USD, 2 USD, etc., hasta un máximo de 32 767 USD.

statement

Cualquier instrucción SELECT, INSERT, UPDATE o DELETE.

## Notas de uso

Las instrucciones preparadas pueden incluir parámetros: valores que se sustituyen en la instrucción cuando se ejecuta. Para incluir parámetros en una instrucción preparada, proporcione una lista de tipos de datos en la instrucción PREPARE y, en la instrucción que se va a preparar, haga referencia a los parámetros por posición utilizando la notación $1, $2, etc. El número máximo de parámetros es 32 767. Cuando ejecute la instrucción, especifique los valores reales de estos parámetros en la instrucción EXECUTE. Para obtener más información, consulta EXECUTE.

Las instrucciones preparadas solo funcionan durante la sesión actual. Cuando la sesión finaliza, se descarta la instrucción preparada; por lo tanto, debe volver a crearla antes de poder usarla nuevamente. Esto también significa que una instrucción preparada no puede utilizarse por varios clientes de base de datos simultáneos. No obstante, cada cliente puede crear su propia

instrucción preparada. La instrucción preparada se puede eliminar manualmente con el comando DEALLOCATE.

Las instrucciones preparadas ofrecen más rendimiento cuando se utiliza una sola sesión para ejecutar una gran cantidad de instrucciones similares. Como se mencionó, para cada ejecución nueva de una instrucción preparada, Amazon Redshift puede revisar el plan de ejecución de la consulta para mejorar el rendimiento a partir de los valores del parámetro especificados. Para examinar el plan de ejecución de la consulta que Amazon Redshift ha elegido para una instrucción EXECUTE específica, utilice el comando EXPLAIN.

Para obtener más información acerca de la planificación de consultas y las estadísticas recopiladas por Amazon Redshift para la optimización de las consultas, consulte el comando ANALYZE.

### Ejemplos

Cree una tabla temporal, prepare la instrucción INSERT y ejecútela:

```
DROP TABLE IF EXISTS prep1;
CREATE TABLE prep1 (c1 int, c2 char(20));
PREPARE prep_insert_plan (int, char)
AS insert into prep1 values ($1, $2);
EXECUTE prep_insert_plan (1, 'one');
EXECUTE prep_insert_plan (2, 'two');
EXECUTE prep_insert_plan (3, 'three');
DEALLOCATE prep_insert_plan;
```

Prepare una instrucción SELECT y ejecútela:

```
PREPARE prep_select_plan (int)
AS select * from prep1 where c1 = $1;
EXECUTE prep_select_plan (2);
EXECUTE prep_select_plan (3);
DEALLOCATE prep_select_plan;
```

### Véase también

DEALLOCATE, EXECUTE

## REFRESH MATERIALIZED VIEW

Actualiza una vista materializada.

Cuando crea una vista materializada, su contenido refleja el estado de la tabla o tablas de bases de datos subyacentes en ese momento. Los datos en la vista materializada permanecen sin cambios, incluso cuando las aplicaciones realizan cambios en los datos de las tablas subyacentes.

Para actualizar los datos en una vista materializada, puede utilizar la instrucción `REFRESH MATERIALIZED VIEW` en cualquier momento. Cuando se utiliza esta instrucción, Amazon Redshift identifica los cambios que han tenido lugar en la tabla o las tablas base, para luego aplicar esos cambios en la vista materializada.

Para obtener más información acerca de las vistas materializadas, consulte Vistas materializadas en Amazon Redshift.

## Sintaxis

```
REFRESH MATERIALIZED VIEW mv_name
```

## Parámetros

mv_name

El nombre de la vista materializada que desea actualizar.

## Notas de uso

Solo el propietario de una vista materializada puede realizar una operación `REFRESH MATERIALIZED VIEW` en esa vista. Además, el propietario debe tener el privilegio SELECT en las tablas base subyacentes para ejecutar correctamente `REFRESH MATERIALIZED VIEW`.

El comando `REFRESH MATERIALIZED VIEW` se ejecuta como una transacción propia. Se sigue la semántica de las transacciones de Amazon Redshift para determinar qué datos de las tablas base son visibles para el comando `REFRESH` o cuándo los cambios realizados por el comando `REFRESH` están visibles para otras transacciones que se ejecutan en Amazon Redshift.

- En las vistas materializadas incrementales, `REFRESH MATERIALIZED VIEW` utiliza solo las filas de la tabla base que ya están confirmadas. Por lo tanto, si la operación de actualización se ejecuta después de una instrucción de lenguaje de manipulación de datos (DML) en la misma transacción, los cambios de esa instrucción DML no son visibles para actualizar.

- Para la actualización completa de una vista materializada, `REFRESH MATERIALIZED VIEW` ve todas las filas de la tabla base visibles para la transacción de actualización, de acuerdo con la semántica de transacción habitual de Amazon Redshift.
- Según el tipo de argumento de entrada, Amazon Redshift sigue admitiendo la actualización incremental para vistas materializadas de las siguientes funciones con tipos de argumentos de entrada específicos: DATE (marca temporal), DATE_PART (fecha, hora, intervalo, tiempo-tz), DATE_TRUNC (marca temporal, intervalo).
- La actualización incremental es compatible con una vista materializada donde la tabla básica se encuentra en un recurso compartido de datos.

Algunas operaciones de Amazon Redshift interactúan con vistas materializadas. Algunas de estas operaciones pueden forzar que una operación `REFRESH MATERIALIZED VIEW` vuelva a calcular completamente la vista materializada aunque la consulta que define la vista materializada solo utilice las características SQL aptas para la actualización incremental. Por ejemplo:

- Las operaciones de vacío en segundo plano pueden bloquearse si las vistas materializadas no se actualizan. Después de un periodo de umbral definido internamente, la operación de vacío puede funcionar. Cuando ocurre esta operación de vacío, cualquier vista materializada dependiente se marca para un nuevo cálculo en la siguiente actualización (incluso si es incremental). Para obtener información acerca de VACUUM, consulte VACUUM. Para obtener más información sobre eventos y cambios de estado, consulte STL_MV_STATE.
- Algunas operaciones iniciadas por un usuario en tablas base fuerzan a la vista materializada a que se vuelva a calcular completamente la próxima vez que se ejecute una operación REFRESH. Ejemplos de tales operaciones son una operación VACUUM invocada manualmente, un cambio de tamaño clásico, una operación ALTER DISTKEY, una operación ALTER SORTKEY y una operación truncada. En algunos casos, las operaciones automáticas también pueden hacer que una vista materializada se vuelva a calcular por completo la próxima vez que se ejecute una operación REFRESH. Por ejemplo, una operación de autovacuum puede provocar un recálculo completo. Para obtener más información sobre eventos y cambios de estado, consulte STL_MV_STATE.

## Actualización incremental para obtener vistas materializadas en un recurso compartido de datos

Amazon Redshift admite la actualización automática e incremental de las vistas materializadas en un recurso compartido de datos de consumidores cuando se comparten las tablas base. La

actualización incremental es una operación en la que Amazon Redshift identifica los cambios en la tabla o tablas base que se produjeron después de la actualización anterior y actualiza solo los registros correspondientes de la vista materializada. Para obtener más información sobre este comportamiento, consulte CREATE MATERIALIZED VIEW.

## Limitaciones para la actualización incremental

En este momento, Amazon Redshift no es compatible con la actualización progresiva para vistas materializadas que se definen con una consulta a través de cualquiera de los siguientes elementos SQL:

- OUTER JOIN (RIGHT, LEFT o FULL).
- Establecer operaciones: UNION, INTERSECT, EXCEPT, MINUS.
- UNION ALL cuando se produce en una subconsulta y una función de agrupación o una cláusula GROUP BY está presente en la consulta.
- Las funciones de agrupación MEDIAN, PERCENTILE_CONT, LISTAGG, STDDEV_SAMP, STDDEV_POP, APPROXIMATE COUNT, APPROXIMATE PERCENTILE y las funciones de agrupación de cada bit.

  > **Note**
  >
  > Se admiten las funciones de agregación COUNT, SUM, MIN, MAX y AVG.

- Funciones de agregación DISTINCT, como DISTINCT COUNT, DISTINCT SUM, etc.
- Funciones de ventana.
- Una consulta que utiliza tablas temporales para la optimización de consultas, como la optimización de subexpresiones comunes.
- subconsultas
- Tablas externas que hacen referencia a los siguientes formatos de la consulta que definen la vista materializada.
  - Delta Lake
  - Hudi

  La actualización incremental es compatible con las vistas materializadas definidas con formatos distintos de los enumerados anteriormente. Para obtener más información, consulte Vistas materializadas de tablas de lagos de datos externos en Amazon Redshift Spectrum.

- Funciones mutables, como funciones de fecha y hora, funciones definidas por el usuario RANDOM y no STABLE.
- Para conocer las limitaciones de la actualización incremental de las integraciones sin ETL, consulte Consideraciones al usar integraciones sin ETL con Amazon Redshift.
- Acceso a tablas desde más de una base de datos.

Para obtener más información sobre las limitaciones de la vista materializada, incluido el efecto de las operaciones en segundo plano, como VACUUM, en las operaciones de actualización de la vista materializada, consulte Notas de uso.

### Ejemplos

El siguiente ejemplo actualiza la vista materializada de `tickets_mv`.

```
REFRESH MATERIALIZED VIEW tickets_mv;
```

## RESET

Restablece el valor de un parámetro de configuración a su valor predeterminado.

Puede restablecer un parámetro especificado o todos los parámetros a la vez. Para establecer un parámetro en un valor específico, use el comando SET. Para mostrar el valor actual de un parámetro, use el comando SHOW.

### Sintaxis

```
RESET { parameter_name | ALL }
```

La siguiente instrucción establece el valor de una variable de contexto de sesión en NULL.

```
RESET { variable_name | ALL }
```

### Parámetros

parameter_name

Nombre del parámetro que se debe restablecer. Consulte Modificación de la configuración del servidor para obtener más información acerca de los parámetros.

ALL

Restablece todos los parámetros de tiempo de ejecución, incluidas todas las variables de contexto de sesión.

variable

El nombre de la variable que restablecer. Si el valor para RESET es una variable de contexto de sesión, Amazon Redshift lo establece en NULL.

### Ejemplos

En el siguiente ejemplo, se restablece el parámetro `query_group` a su valor predeterminado:

```
reset query_group;
```

En el siguiente ejemplo, se restablecen todos los parámetros de tiempo de ejecución a sus valores predeterminados.

```
reset all;
```

En el ejemplo siguiente, se restablece la variable de contexto.

```
RESET app_context.user_id;
```

## REVOKE

Quita los permisos de acceso, como los permisos para crear, eliminar o actualizar tablas, a un usuario o rol.

Solo puede dar permisos GRANT o REVOKE USAGE en un esquema externo a los usuarios de la base de datos y los roles con la sintaxis ON SCHEMA. Cuando use ON EXTERNAL SCHEMA con AWS Lake Formation, solo puede usar GRANT y REVOKE para conceder y revocar permisos a un rol de AWS Identity and Access Management (IAM). Para obtener una lista de los permisos, consulte la sintaxis.

Para procedimientos almacenados, se concede el permiso USAGE ON LANGUAGE `plpgsql` a PUBLIC de manera predeterminada. El permiso EXECUTE ON PROCEDURE se concede solo al propietario y a los superusuarios de manera predeterminada.

Especifique en el comando REVOKE los permisos que desea eliminar. Para conceder permisos, utilice el comando GRANT.

## Sintaxis

```
REVOKE [ GRANT OPTION FOR ]
{ { SELECT | INSERT | UPDATE | DELETE | DROP | REFERENCES | ALTER | TRUNCATE } [,...] |
 ALL [ PRIVILEGES ] }
ON { [ TABLE ] table_name [, ...] | ALL TABLES IN SCHEMA schema_name [, ...] }
FROM { username | ROLE role_name | GROUP group_name | PUBLIC } [, ...]
[ RESTRICT ]

REVOKE [ GRANT OPTION FOR ]
{ { CREATE | TEMPORARY | TEMP | ALTER } [,...] | ALL [ PRIVILEGES ] }
ON DATABASE db_name [, ...]
FROM { username | ROLE role_name | GROUP group_name | PUBLIC } [, ...]
[ RESTRICT ]

REVOKE [ GRANT OPTION FOR ]
{ { CREATE | USAGE | ALTER | DROP } [,...] | ALL [ PRIVILEGES ] }
ON SCHEMA schema_name [, ...]
FROM { username | ROLE role_name | GROUP group_name | PUBLIC } [, ...]
[ RESTRICT ]

REVOKE [ GRANT OPTION FOR ]
EXECUTE
    ON FUNCTION function_name ( [ [ argname ] argtype [, ...] ] ) [, ...]
    FROM { username | ROLE role_name | GROUP group_name | PUBLIC } [, ...]
[ RESTRICT ]

REVOKE [ GRANT OPTION FOR ]
{ { EXECUTE } [,...] | ALL [ PRIVILEGES ] }
    ON PROCEDURE procedure_name ( [ [ argname ] argtype [, ...] ] ) [, ...]
    FROM { username | ROLE role_name | GROUP group_name | PUBLIC } [, ...]
[ RESTRICT ]

REVOKE [ GRANT OPTION FOR ]
USAGE
    ON LANGUAGE language_name [, ...]
    FROM { username | ROLE role_name | GROUP group_name | PUBLIC } [, ...]
[ RESTRICT ]

REVOKE [GRANT OPTION FOR]
```

```
{ { ALTER | DROP} [,...] | ALL [ PRIVILEGES ] }
    ON COPY JOB job_name [,...]
    FROM { username | ROLE role_name | GROUP group_name | PUBLIC } [, ...]
```

Revocación de permisos de nivel de columna para tablas

A continuación, se muestra la sintaxis de los permisos de nivel de columna en tablas y vistas de Amazon Redshift.

```
REVOKE { { SELECT | UPDATE } ( column_name [, ...] ) [, ...] | ALL [ PRIVILEGES ]
 ( column_name [,...] ) }
     ON { [ TABLE ] table_name [, ...] }
     FROM { username | ROLE role_name | GROUP group_name | PUBLIC } [, ...]
     [ RESTRICT ]
```

Revocación de los permisos ASSUMEROLE

A continuación, se muestra la sintaxis para revocar el permiso ASSUMEROLE a los usuarios y grupos con un rol especificado.

```
REVOKE ASSUMEROLE
    ON { 'iam_role' [, ...]  | default | ALL }
    FROM { user_name | ROLE role_name | GROUP group_name | PUBLIC } [, ...]
    FOR { ALL | COPY | UNLOAD | EXTERNAL FUNCTION | CREATE MODEL }
```

Revocación de permisos para Redshift Spectrum para Lake Formation

A continuación, se muestra la sintaxis para la integración de Redshift Spectrum con Lake Formation.

```
REVOKE [ GRANT OPTION FOR ]
{ SELECT | ALL [ PRIVILEGES ] } ( column_list )
    ON EXTERNAL TABLE schema_name.table_name
    FROM { IAM_ROLE iam_role } [, ...]

REVOKE [ GRANT OPTION FOR ]
{ { SELECT | ALTER | DROP | DELETE | INSERT }  [, ...] | ALL [ PRIVILEGES ] }
    ON EXTERNAL TABLE schema_name.table_name [, ...]
    FROM { { IAM_ROLE iam_role } [, ...] | PUBLIC }

REVOKE [ GRANT OPTION FOR ]
{ { CREATE | ALTER | DROP }  [, ...] | ALL [ PRIVILEGES ] }
```

```
    ON EXTERNAL SCHEMA schema_name [, ...]
    FROM { IAM_ROLE iam_role } [, ...]
```

Revocación de permisos de recurso compartido de datos

Permisos de recursos compartidos de datos del lado del productor

A continuación, se muestra la sintaxis para utilizar REVOKE para quitar permisos ALTER o SHARE a un usuario o rol. El usuario cuyos permisos se han revocado ya no puede modificar el recurso compartido de datos ni conceder el uso a un consumidor.

```
REVOKE { ALTER | SHARE } ON DATASHARE datashare_name
 FROM { username [ WITH GRANT OPTION ] | ROLE role_name | GROUP group_name | PUBLIC }
 [, ...]
```

A continuación, se muestra la sintaxis para utilizar REVOKE para eliminar el acceso de un consumidor a un recurso compartido de datos.

```
REVOKE USAGE
 ON DATASHARE datashare_name
 FROM NAMESPACE 'namespaceGUID' [, ...] | ACCOUNT 'accountnumber' [ VIA DATA CATALOG ]
 [, ...]
```

A continuación, se ofrece un ejemplo de cómo revocar el uso de un recurso compartido de datos de una cuenta de Lake Formation.

```
REVOKE USAGE ON DATASHARE salesshare FROM ACCOUNT '123456789012' VIA DATA CATALOG;
```

Permisos de recursos compartidos de datos del lado del consumidor

A continuación, se muestra la sintaxis de REVOKE para los permisos de uso compartido de datos en una base de datos o un esquema específicos creados a partir de un datashare. La revocación del permiso de uso de una base de datos creada con la cláusula WITH PERMISSIONS no revoca ningún permiso adicional que haya concedido a un usuario o rol, ni siquiera los permisos de nivel de objeto concedidos a los objetos subyacentes. Si vuelve a conceder el permiso de uso a ese usuario o rol, conservarán todos los permisos adicionales que tenían antes de que revocara el uso.

```
REVOKE USAGE ON { DATABASE shared_database_name [, ...] | SCHEMA shared_schema}
```

```
 FROM { username | ROLE role_name | GROUP group_name | PUBLIC } [, ...]
```

Revocación de los permisos acotados

Los permisos limitados le permiten conceder permisos a un usuario o rol en todos los objetos de un tipo dentro de una base de datos o un esquema. Los usuarios y roles con permisos limitados tienen los permisos especificados en todos los objetos actuales y futuros de la base de datos o del esquema.

Puede ver el alcance de los permisos limitados en el nivel de base de datos en SVV_DATABASE_PRIVILEGES. Puede ver el alcance de los permisos limitados en el nivel de esquema en SVV_SCHEMA_PRIVILEGES.

Para obtener más información sobre los permisos acotados, consulte Permisos acotados.

A continuación, se muestra la sintaxis para revocar permisos acotados a usuarios y roles.

```
REVOKE [ GRANT OPTION ]
{ CREATE | USAGE | ALTER | DROP } [,...] | ALL [ PRIVILEGES ] }
FOR SCHEMAS IN
DATABASE db_name
FROM { username | ROLE role_name } [, ...]

REVOKE [ GRANT OPTION ]
{ { SELECT | INSERT | UPDATE | DELETE | DROP | ALTER | TRUNCATE | REFERENCES }
 [, ...] } | ALL [PRIVILEGES] } }
FOR TABLES IN
{ SCHEMA schema_name [ DATABASE db_name ] | DATABASE db_name }
FROM { username | ROLE role_name } [, ...]

REVOKE [ GRANT OPTION ] { EXECUTE | ALL [ PRIVILEGES ] }
FOR FUNCTIONS IN
{ SCHEMA schema_name [DATABASE db_name ] | DATABASE db_name }
FROM { username | ROLE role_name } [, ...]

REVOKE [ GRANT OPTION ] { EXECUTE | ALL [ PRIVILEGES ] }
FOR PROCEDURES IN
{ SCHEMA schema_name [DATABASE db_name ] | DATABASE db_name }
FROM { username | ROLE role_name } [, ...]

REVOKE [ GRANT OPTION ] USAGE
FOR LANGUAGES IN
DATABASE db_name
```

```
FROM { username | ROLE role_name } [, ...]

REVOKE [GRANT_OPTION]
{ { CREATE | ALTER | DROP} [,...] | ALL [ PRIVILEGES ] }
FOR COPY JOBS
IN DATABASE db_name
FROM { username [ WITH GRANT OPTION ] | ROLE role_name } [, ...]
```

Tenga en cuenta que los permisos limitados no distinguen entre los permisos de las funciones y los de los procedimientos. Por ejemplo, la siguiente instrucción revoca los permisos EXECUTE tanto para las funciones como para los procedimientos de bob en el esquema Sales_schema.

```
REVOKE EXECUTE FOR FUNCTIONS IN SCHEMA Sales_schema FROM bob;
```

Revocación de permisos de machine learning

A continuación, se muestra la sintaxis para los permisos de modelos de machine learning en Amazon Redshift.

```
REVOKE [ GRANT OPTION FOR ]
    CREATE MODEL FROM { username | ROLE role_name | GROUP group_name | PUBLIC } [, ...]
    [ RESTRICT ]

REVOKE [ GRANT OPTION FOR ]
    { EXECUTE | ALL [ PRIVILEGES ] }
    ON MODEL model_name [, ...]

    FROM { username | ROLE role_name | GROUP group_name | PUBLIC } [, ...]
    [ RESTRICT ]
```

Revocación de los permisos de rol

A continuación, se muestra la sintaxis para revocar permisos de rol en Amazon Redshift.

```
REVOKE [ ADMIN OPTION FOR ] { ROLE role_name } [, ...] FROM { user_name } [, ...]
```

```
REVOKE { ROLE role_name } [, ...] FROM { ROLE role_name } [, ...]
```

A continuación, se muestra la sintaxis para revocar permisos de sistema a los roles en Amazon Redshift.

```
REVOKE
  {
    { CREATE USER | DROP USER | ALTER USER |
    CREATE SCHEMA | DROP SCHEMA |
    ALTER DEFAULT PRIVILEGES |
    ACCESS CATALOG |
    CREATE TABLE | DROP TABLE | ALTER TABLE |
    CREATE OR REPLACE FUNCTION | CREATE OR REPLACE EXTERNAL FUNCTION |
    DROP FUNCTION |
    CREATE OR REPLACE PROCEDURE | DROP PROCEDURE |
    CREATE OR REPLACE VIEW | DROP VIEW |
    CREATE MODEL | DROP MODEL |
    CREATE DATASHARE | ALTER DATASHARE | DROP DATASHARE |
    CREATE LIBRARY | DROP LIBRARY |
    CREATE ROLE | DROP ROLE
    TRUNCATE TABLE
    VACUUM | ANALYZE | CANCEL }[, ...]
  }
  | { ALL [ PRIVILEGES ] }
FROM { ROLE role_name } [, ...]
```

Revocación de permisos EXPLAIN para filtros de políticas de seguridad de nivel de fila

Co la siguiente sintaxis, se puede revocar permisos para explicar los filtros de la política de seguridad de la fila de una consulta en el plan EXPLAIN. Puede revocar el privilegio mediante la instrucción REVOKE.

```
REVOKE EXPLAIN RLS FROM ROLE rolename
```

Con la siguiente sintaxis, se puede conceder permisos para omitir las políticas de seguridad de la fila para una consulta.

```
REVOKE IGNORE RLS FROM ROLE rolename
```

A continuación, se muestra la sintaxis para revocar permisos de la política de seguridad de la fila especificada.

```
REVOKE SELECT ON [ TABLE ] table_name [, ...]
            FROM RLS POLICY policy_name [, ...]
```

## Parámetros

GRANT OPTION FOR

Revoca solo la opción de conceder un permiso especificado a otros usuarios y no revoca el permiso en sí. No puede revocar GRANT OPTION de un grupo o de PUBLIC.

SELECT

Revoca el privilegio para seleccionar datos de una tabla o vista a través de una instrucción SELECT.

INSERT

Revoca el permiso para cargar datos en una tabla a través de una instrucción INSERT o una instrucción COPY.

UPDATE

Revoca el privilegio para actualizar una columna de la tabla a través de una instrucción UPDATE.

DELETE

Revoca el permiso para eliminar una fila de datos de una tabla.

REFERENCES

Revoca el permiso para crear una restricción de clave externa. Debe revocar este permiso en la tabla a la que se hace referencia y en la tabla de referencia.

TRUNCATE

Revoca el permiso para truncar una tabla. Sin este permiso, solo el propietario de una tabla o un superusuario pueden truncarla. Para obtener más información sobre el comando TRUNCATE, consulte the section called “TRUNCATE”.

ALL [ PRIVILEGES ]

Revoca todos los permisos disponibles a la vez del usuario o grupo especificado. La palabra clave PRIVILEGES es opcional.

> **Note**
>
> Amazon Redshift no admite los permisos RULE y TRIGGER. Para obtener más información, consulte Características no compatibles de PostgreSQL.

ALTER

Según el objeto de base de datos, revoca los siguientes permisos del usuario o grupo de usuarios:

- En el caso de las tablas, ALTER revoca el permiso para modificar una tabla o una vista. Para obtener más información, consulte ALTER TABLE.
- En el caso de las bases de datos, ALTER revoca el permiso para modificar una base de datos. Para obtener más información, consulte ALTER DATABASE.
- En el caso de los esquemas, ALTER concede revocaciones para modificar un esquema. Para obtener más información, consulte ALTER SCHEMA.
- En el caso de las tablas externas, ALTER revoca el permiso para alterar una tabla en un AWS Glue Data Catalog que esté habilitado para Lake Formation. Este permiso solo se aplica cuando se utiliza Lake Formation.

DROP

Según el objeto de base de datos, revoca los siguientes permisos del usuario o rol:

- En el caso de las tablas, DROP revoca el permiso para eliminar una tabla o una vista. Para obtener más información, consulte DROP TABLE.
- En el caso de las bases de datos, DROP revoca el permiso para eliminar una base de datos. Para obtener más información, consulte DROP DATABASE.
- En el caso de los esquemas, DROP revoca el permiso para eliminar un esquema. Para obtener más información, consulte DROP SCHEMA.

ASSUMEROLE

Revoca el permiso para ejecutar comandos COPY, UNLOAD, EXTERNAL FUNCTION o CREATE MODEL de usuarios, roles o grupos con un rol especificado.

ON [ TABLE ] table_name

Revoca los permisos especificados en una tabla o vista. La palabra clave TABLE es opcional.

ON ALL TABLES IN SCHEMA schema_name (nombre_de_esquema)

Revoca los permisos especificados en todas las tablas en el esquema al que se hace referencia.

( column_name [,...] ) ON TABLE table_name

Revoca los permisos especificados de usuarios, grupos o PUBLIC en las columnas especificadas de la tabla o la vista de Amazon Redshift.

( column_list ) ON EXTERNAL TABLE schema_name.table_name

Revoca los permisos especificados de un rol de IAM en las columnas especificadas de la tabla de Lake Formation en el esquema de la referencia.

ON EXTERNAL TABLE schema_name.table_name

Revoca los permisos especificados de un rol de IAM en las tablas de Lake Formation especificadas en el esquema de la referencia.

ON EXTERNAL SCHEMA schema_name

Revoca los permisos especificados de un rol de IAM en el esquema al que se hace referencia.

FROM IAM_ROLE iam_role

Indica el rol de IAM que pierde los permisos.

ROLE role_name

Revoca los permisos del rol especificado.

GROUP group_name

Revoca los permisos del grupo de usuarios especificado.

PUBLIC

Revoca los permisos especificados de todos los usuarios. PUBLIC representa un grupo que siempre incluye a todos los usuarios. Los permisos de un usuario individual constan de la suma de permisos concedidos a PUBLIC, los permisos concedidos a cualquier grupo al que pertenezca el usuario y los permisos concedidos al usuario de manera individual.

La revocación de PUBLIC de una tabla externa de Lake Formation da lugar a la revocación del permiso del grupo everyone de Lake Formation.

CREATE

Según el objeto de base de datos, revoca los siguientes permiso del usuario o grupo de usuarios:

- Para las bases de datos, usar la cláusula CREATE para REVOKE previene que los usuarios creen esquemas dentro de la base de datos.
- Para los esquemas, usar la cláusula CREATE para REVOKE previene que los usuarios creen objetos dentro de un esquema. Para cambiar el nombre de un objeto, el usuario debe tener el permiso CREATE y ser propietario del objeto cuyo nombre va a cambiarse.

> **Note**
>
> De manera predeterminada, todos los usuarios tienen permisos CREATE y USAGE en el esquema PUBLIC.

TEMPORARY | TEMP

Revoca el permiso para crear tablas temporales en la base de datos especificada.

> **Note**
>
> Por defecto, los usuarios reciben permisos para crear tablas temporales con su membresía automática en el grupo PUBLIC. Para eliminar el permiso de cualquier usuario para crear tablas temporales, revoque el permiso TEMP del grupo PUBLIC y, después, conceda explícitamente el permiso para crear tablas temporales solo a usuarios o grupos de usuarios específicos.

ON DATABASE db_name

Revoca los permisos en la base de datos especificada.

USAGE

Revoca los permisos USAGE en objetos de un esquema específico, lo que provoca que los usuarios no puedan acceder a estos objetos. Las acciones específicas en estos objetos deben revocarse por separado (como el permiso EXECUTE en las funciones).

> **Note**
>
> De manera predeterminada, todos los usuarios tienen permisos CREATE y USAGE en el esquema PUBLIC.

ON SCHEMA schema_name

Revoca los permisos en el esquema especificado. Puede usar permisos de esquema para controlar la creación de tablas. El permiso CREATE para una base de datos solo controla la creación de esquemas.

RESTRICT

Revoca solo los permisos que el usuario concedió directamente. Este es el comportamiento predeterminado.

EXECUTE ON PROCEDURE procedure_name

Revoca el permiso EXECUTE en un procedimiento almacenado específico. Debido a que los nombres de procedimientos almacenados no se pueden sobrecargar, debe incluir la lista de argumentos para el procedimiento. Para obtener más información, consulte Nomenclatura de los procedimientos almacenados.

EXECUTE ON ALL PROCEDURES IN SCHEMA procedure_name

Revoca los permisos especificados en todos los procedimientos en el esquema al que se hace referencia.

USAGE ON LANGUAGE language_name

Revoca el permiso USAGE en un lenguaje. Para Funciones definidas por el usuario (UDF) utilice `plpythonu`. Para las UDF de SQL, use `sql`. Para procedimientos almacenados, utilice `plpgsql`.

Para crear una UDF, debe tener permiso para usar el lenguaje de SQL o `plpythonu` (Python). De manera predeterminada, se concede el permiso USAGE ON LANGUAGE SQL a PUBLIC. No obstante, debe conceder explícitamente el permiso de USAGE ON LANGUAGE PLPYTHONU a usuarios o grupos específicos.

Para revocar los permisos de uso de SQL, primero revoque el uso de PUBLIC. A continuación, conceda los permisos de uso de SQL a solo los usuarios o grupos que tienen permiso para crear UDF de SQL. En el siguiente ejemplo, se revocan los permisos de uso de SQL de PUBLIC y después se conceden permisos de uso al grupo de usuarios `udf_devs`.

```
revoke usage on language sql from PUBLIC;
grant usage on language sql to group udf_devs;
```

Para obtener más información, consulte Seguridad y permisos de UDF.

Para revocar el uso para procedimientos almacenados, primero revoque el uso de PUBLIC. A continuación, conceda los permisos de uso en `plpgsql` solo a los usuarios o grupos que tienen permiso para crear procedimientos almacenados. Para obtener más información, consulte Seguridad y privilegios para procedimientos almacenados .

ON COPY JOB job_name

Revoca los permisos especificados en un trabajo de copia.

FOR { ALL | COPY | UNLOAD | EXTERNAL FUNCTION | CREATE MODEL } [, ...]

Especifica el comando SQL para el cual se revoca el permiso. Puede especificar ALL para revocar el permiso en las instrucciones COPY, UNLOAD, EXTERNAL FUNCTION y CREATE MODEL. Esta cláusula se aplica solo a la revocación del permiso ASSUMEROLE.

ALTER

Revoca el permisos ALTER para usuarios o grupos de usuarios, lo que permite modificar un recurso compartido de datos a quienes no lo poseen. Este permiso es necesario para agregar o quitar objetos de un recurso compartido de datos o para establecer la propiedad PUBLICACCESSIBLE. Para obtener más información, consulte ALTER DATASHARE.

SHARE

Revoca permisos para usuarios y grupos de usuarios que sirven para agregar consumidores a un recurso compartido de datos. Es necesario revocar este permiso para que un consumidor concreto deje de acceder al recurso compartido de datos desde sus clústeres.

ON DATASHARE datashare_name

Concede los permisos especificados en el recurso compartido de datos al que se hace referencia.

FROM username

Indica el usuario que pierde los permisos.

FROM GROUP group_name

Indica el grupo de usuarios que pierde los permisos.

WITH GRANT OPTION

Indica que el usuario que pierde los permisos puede, a su vez, revocar los mismos permisos a otros usuarios. No se puede revocar WITH GRANT OPTION para un grupo o para PUBLIC.

USAGE

Cuando se revoca USAGE para una cuenta consumidora o un espacio de nombres dentro de la misma cuenta, estos no pueden acceder al datashare y a los objetos del datashare para lectura solamente.

La revocación del permiso USAGE revoca el acceso de los consumidores a un recurso compartido de datos.

FROM NAMESPACE 'clusternamespace GUID'

Indica el espacio de nombres en la misma cuenta que tiene consumidores que pierden los permisos en el recurso compartido de datos. Los espacios de nombres utilizan un identificador único global (GUID) alfanumérico de 128 bits.

FROM ACCOUNT 'número_de_cuenta' [ VIA DATA CATALOG ]

Indica el número de otra cuenta que tiene consumidores que pierden los permisos en el recurso compartido de datos. Si especifica “VIA DATA CATALOG”, se indica que revoca el uso del recurso compartido de datos de una cuenta de Lake Formation. Omitir el número de cuenta significa que se revoca de la cuenta que es propietaria del clúster.

ON DATABASE shared_database_name**>** [, ...]

Revoca los permisos de uso especificados en la base de datos concreta que se creó en el recurso compartido de datos indicado.

ON SCHEMA shared_schema

Revoca los permisos especificados en el esquema concreto que se creó en el recurso compartido de datos indicado.

FOR { SCHEMAS | TABLES | FUNCTIONS | PROCEDURES | LANGUAGES | COPY JOBS} IN

Especifica los objetos de la base de datos a los que se va a revocar el permiso. Los parámetros que siguen a IN definen el alcance del permiso revocado.

CREATE MODEL

Revoca el permiso CREATE MODEL para crear modelos de machine learning en la base de datos especificada.

ON MODEL model_name

Revoca el permiso EXECUTE para un modelo específico.

ACCESS CATALOG

Revoca el permiso para ver los metadatos relevantes de los objetos a los que tiene acceso el rol.

[ ADMIN OPTION FOR ] { role } [, …]

Rol que se revoca a un usuario especificado que tiene WITH ADMIN OPTION.

FROM { role } [, …]

Rol al que se le revoca el rol especificado.

## Notas de uso

Para obtener más información acerca de las notas de uso de REVOKE, consulte the section called “Notas de uso”.

## Ejemplos

Para ver ejemplos de cómo utilizar REVOKE, consulte the section called “Ejemplos”.

## Notas de uso

Para revocar los privilegios de un objeto, debe cumplir con uno de los siguientes criterios:

- Ser el propietario del objeto.
- Ser un superusuario.
- Tener un privilegio concedido para ese objeto y privilegio.

  Por ejemplo, el siguiente comando permite al usuario HR realizar comandos SELECT en la tabla de empleados y conceder y revocar el mismo privilegio para otros usuarios:

  ```
  grant select on table employees to HR with grant option;
  ```

  HR no puede revocar privilegios para ninguna operación que no sea SELECT o ninguna tabla que no sea la de empleados.

Los superusuarios pueden obtener acceso a todos los objetos independientemente de los comandos GRANT y REVOKE que establecen privilegios de objeto.

PUBLIC representa un grupo que siempre incluye a todos los usuarios. De manera predeterminada, todos los miembros de PUBLIC, tienen privilegios CREATE y USAGE en el esquema PUBLIC. Para restringir los permisos de los usuarios en el esquema PUBLIC, primero debe revocar todos los permisos de PUBLIC en el esquema PUBLIC y después conceder privilegios a usuarios o grupos específicos. En el siguiente ejemplo, se controlan los privilegios de creación de tablas en el esquema PUBLIC.

```
revoke create on schema public from public;
```

Para revocar privilegios de una tabla de Lake Formation, el rol de IAM asociado al esquema externo de la tabla debe tener permiso para revocar privilegios a la tabla externa. El siguiente ejemplo crea un esquema externo con un rol IAM `myGrantor` asociado. El rol IAM `myGrantor` tiene el permiso para retirar permisos a otros. El comando REVOKE utiliza los permisos del rol `myGrantor` IAM que está asociado al esquema externo para retirar permisos al rol IAM `myGrantee`.

```
create external schema mySchema
from data catalog
database 'spectrum_db'
iam_role 'arn:aws:iam::123456789012:role/myGrantor'
create external database if not exists;
```

```
revoke select
on external table mySchema.mytable
from iam_role 'arn:aws:iam::123456789012:role/myGrantee';
```

> **Note**
>
> Si el rol de IAM también tiene el permiso `ALL` en un AWS Glue Data Catalog que está habilitado para Lake Formation, no se revoca el permiso `ALL`. Solo el permiso `SELECT` se retira. Puede ver los permisos de Lake Formation en la consola de dicho servicio.

Notas de uso para revocar el permiso ASSUMEROLE

Las siguientes notas de uso se aplican para revocar el privilegio ASSUMEROLE en Amazon Redshift.

Solo un superusuario de base de datos puede revocar el privilegio ASSUMEROLE para usuarios y grupos. Un superusuario siempre conserva el privilegio ASSUMEROLE.

Para habilitar el uso del privilegio ASSUMEROLE para usuarios y grupos, un superusuario ejecuta una vez la siguiente instrucción en el clúster. Antes de conceder el privilegio ASSUMEROLE a usuarios y grupos, un superusuario debe ejecutar una vez la siguiente instrucción en el clúster.

```
revoke assumerole on all from public for all;
```

### Notas de uso para revocar permisos de machine learning

No puede conceder ni revocar directamente permisos relacionados con una función de ML. Una función de ML pertenece a un modelo de ML y los permisos se controlan mediante el modelo. En su lugar, puede revocar los permisos relacionados con el modelo de ML. En el siguiente ejemplo, se muestra cómo revocar el permiso de ejecución a todos los usuarios asociados al modelo `customer_churn`.

```
REVOKE EXECUTE ON MODEL customer_churn FROM PUBLIC;
```

También puede revocar todos los permisos a un usuario para el modelo de ML `customer_churn`.

```
REVOKE ALL on MODEL customer_churn FROM ml_user;
```

Se producirá un error en la revocación del permiso `EXECUTE` relacionado con una función de ML si existe una función de ML en el esquema, aunque dicha función de ML ya disponga del permiso `EXECUTE` mediante `GRANT EXECUTE ON MODEL`. Recomendamos utilizar un esquema independiente cuando utilice el comando `CREATE MODEL` para mantener las funciones de ML en un esquema independiente por sí mismas. En el siguiente ejemplo, se muestra cómo hacerlo.

```
CREATE MODEL ml_schema.customer_churn
FROM customer_data
TARGET churn
FUNCTION ml_schema.customer_churn_prediction
IAM_ROLE default
SETTINGS (
 S3_BUCKET 'amzn-s3-demo-bucket'
);
```

## Ejemplos

En el siguiente ejemplo, se revocan privilegios INSERT en la tabla SALES del grupo de usuarios GUESTS. Este comando previene que los miembros de GUESTS puedan cargar datos en la tabla SALES con el comando INSERT.

```
revoke insert on table sales from group guests;
```

En el siguiente ejemplo, se le revoca el privilegio SELECT en todas las tablas del esquema QA_TICKIT al usuario `fred`.

```
revoke select on all tables in schema qa_tickit from fred;
```

En el siguiente ejemplo, se le revoca el privilegio para seleccionar desde una vista al usuario `bobr`.

```
revoke select on table eventview from bobr;
```

En el siguiente ejemplo, se le revoca el privilegio para crear tablas temporales en la base de datos TICKIT a todos los usuarios.

```
revoke temporary on database tickit from public;
```

En el siguiente ejemplo, se revoca el privilegio SELECT en las columnas `cust_name` y `cust_phone` de la tabla `cust_profile` del usuario `user1`.

```
revoke select(cust_name, cust_phone) on cust_profile from user1;
```

En el siguiente ejemplo, se revoca el privilegio SELECT en las columnas `cust_name` y `cust_phone` y el privilegio UPDATE en la columna `cust_contact_preference` de la tabla `cust_profile` del grupo `sales_group`.

```
revoke select(cust_name, cust_phone), update(cust_contact_preference) on cust_profile
 from group sales_group;
```

En el siguiente ejemplo, se muestra el uso de la palabra clave ALL para revocar los privilegios SELECT y UPDATE de las tres columnas de la tabla `cust_profile` del grupo `sales_admin`.

```
revoke ALL(cust_name, cust_phone,cust_contact_preference) on cust_profile from group
 sales_admin;
```

En el siguiente ejemplo, se revoca el privilegio SELECT de la columna `cust_name` de la vista `cust_profile_vw` del usuario `user2`.

```
revoke select(cust_name) on cust_profile_vw from user2;
```

Ejemplos de revocación del permiso USAGE a bases de datos creadas a partir de recursos compartidos de datos

En el ejemplo siguiente se revoca el acceso al recurso compartido de datos `salesshare` desde el espacio de nombres `13b8833d-17c6-4f16-8fe4-1a018f5ed00d`.

```
REVOKE USAGE ON DATASHARE salesshare FROM NAMESPACE
 '13b8833d-17c6-4f16-8fe4-1a018f5ed00d';
```

En el siguiente ejemplo se revoca el permiso USAGE en `sales_db` a `Bob`.

```
REVOKE USAGE ON DATABASE sales_db FROM Bob;
```

El siguiente es un ejemplo de permiso REVOKE USAGE en `sales_schema` para el `Analyst_role`.

```
REVOKE USAGE ON SCHEMA sales_schema FROM ROLE Analyst_role;
```

Ejemplos de revocación de permisos acotados

El siguiente ejemplo revoca el uso de todos los esquemas actuales y futuros de la base de datos `Sales_db` al rol `Sales`.

```
REVOKE USAGE FOR SCHEMAS IN DATABASE Sales_db FROM ROLE Sales;
```

El siguiente ejemplo revoca la posibilidad de conceder el permiso SELECT para todas las tablas actuales y futuras de la base de datos `Sales_db` al usuario `alice`. `alice` conserva el acceso a todas las tablas en `Sales_db`.

```
REVOKE GRANT OPTION SELECT FOR TABLES IN DATABASE Sales_db FROM alice;
```

En el siguiente ejemplo, se revoca el permiso EXECUTE para las funciones del esquema `Sales_schema` al usuario `bob`.

```
REVOKE EXECUTE FOR FUNCTIONS IN SCHEMA Sales_schema FROM bob;
```

En el siguiente ejemplo se revocan todos los permisos para todas las tablas del esquema `ShareSchema` de la base de datos `ShareDb` al rol `Sales`. Al especificar el esquema, también puede indicar la base de datos del esquema mediante el formato de dos partes `database.schema`.

```
REVOKE ALL FOR TABLES IN SCHEMA ShareDb.ShareSchema FROM ROLE Sales;
```

El siguiente ejemplo es el mismo que el anterior. Puede especificar la base de datos del esquema mediante la palabra clave `DATABASE` en lugar de utilizar un formato de dos partes.

```
REVOKE ALL FOR TABLES IN SCHEMA ShareSchema DATABASE ShareDb FROM ROLE Sales;
```

Ejemplos de revocación del privilegio ASSUMEROLE

A continuación, se muestran ejemplos de revocación del privilegio ASSUMEROLE.

Un superusuario debe habilitar el uso del privilegio ASSUMEROLE para usuarios y grupos al ejecutar una sola vez la siguiente instrucción en el clúster:

```
revoke assumerole on all from public for all;
```

La siguiente instrucción revoca el privilegio ASSUMEROLE del usuario reg_user1 en todos los roles para todas las operaciones.

```
revoke assumerole on all from reg_user1 for all;
```

Ejemplos de revocación del privilegio ROLE

El siguiente ejemplo revoca sample_role1 a sample_role2.

```
CREATE ROLE sample_role2;
GRANT ROLE sample_role1 TO ROLE sample_role2;
REVOKE ROLE sample_role1 FROM ROLE sample_role2;
```

El siguiente ejemplo revoca los privilegios del sistema a user1.

```
GRANT ROLE sys:DBA TO user1;
REVOKE ROLE sys:DBA FROM user1;
```

El siguiente ejemplo revoca sample_role1 y sample_role2 a user1.

```
CREATE ROLE sample_role1;
CREATE ROLE sample_role2;
GRANT ROLE sample_role1, ROLE sample_role2 TO user1;
REVOKE ROLE sample_role1, ROLE sample_role2 FROM user1;
```

El siguiente ejemplo revoca sample_role2 con ADMIN OPTION a user1.

```
GRANT ROLE sample_role2 TO user1 WITH ADMIN OPTION;
```

```
REVOKE ADMIN OPTION FOR ROLE sample_role2 FROM user1;
REVOKE ROLE sample_role2 FROM user1;
```

El siguiente ejemplo revoca sample_role1 y sample_role2 a sample_role5.

```
CREATE ROLE sample_role5;
GRANT ROLE sample_role1, ROLE sample_role2 TO ROLE sample_role5;
REVOKE ROLE sample_role1, ROLE sample_role2 FROM ROLE sample_role5;
```

El siguiente ejemplo revoca los privilegios del sistema CREATE SCHEMA y DROP SCHEMA a sample_role1.

```
GRANT CREATE SCHEMA, DROP SCHEMA TO ROLE sample_role1;
REVOKE CREATE SCHEMA, DROP SCHEMA FROM ROLE sample_role1;
```

# ROLLBACK

Detiene la transacción actual y descarta todas las actualizaciones realizadas por esa transacción.

Este comando lleva a cabo la misma función que el comando ABORT.

## Sintaxis

```
ROLLBACK [ WORK | TRANSACTION ]
```

## Parámetros

WORK

Palabra clave opcional. Esta palabra clave no se admite en un proceso almacenado.

TRANSACTION

Palabra clave opcional. WORK y TRANSACTION son sinónimos. Ninguna de ellas se admite en un proceso almacenado.

Para más información sobre el uso de ROLLBACK en un proceso almacenado, vea Administración de transacciones.

## Ejemplo

En el siguiente ejemplo, se crea una tabla que inicia una transacción donde los datos se insertan en la tabla. El comando ROLLBACK revierte la inserción de datos para dejar la tabla vacía.

El siguiente comando crea una tabla de ejemplo denominada MOVIE_GROSS:

```
create table movie_gross( name varchar(30), gross bigint );
```

El siguiente conjunto de comandos inicia una transacción en donde se insertan dos filas de datos en la tabla:

```
begin;

insert into movie_gross values ( 'Raiders of the Lost Ark', 23400000);

insert into movie_gross values ( 'Star Wars', 10000000 );
```

A continuación, el siguiente comando selecciona los datos de la tabla para mostrar que se insertaron correctamente:

```
select * from movie_gross;
```

En el resultado del comando se muestra que ambas filas se insertaron correctamente:

```
name           |  gross
-------------------------+----------
Raiders of the Lost Ark | 23400000
Star Wars               | 10000000
(2 rows)
```

Este comando ahora revierte los cambios de datos para que vuelvan al estado de inicio de la transacción:

```
rollback;
```

Si selecciona los datos de la tabla, ahora se muestra una tabla vacía:

```
select * from movie_gross;
```

```
name | gross
------+-------
(0 rows)
```

# SELECT

Devuelve filas de tablas, vistas y funciones definidas por el usuario.

> **Note**
>
> El tamaño máximo de una instrucción SQL es de 16 MB.

## Sintaxis

```
[ WITH with_subquery [, ...] ]
SELECT
[ TOP number | [ ALL | DISTINCT ]
* | expression [ AS output_name ] [, ...] ]
[ EXCLUDE column_list ]
[ FROM table_reference [, ...] ]
[ WHERE condition ]
[ [ START WITH expression ] CONNECT BY expression ]
[ GROUP BY ALL | expression [, ...] ]
[ HAVING condition ]
[ QUALIFY condition ]
[ { UNION | ALL | INTERSECT | EXCEPT | MINUS } query ]
[ ORDER BY expression [ ASC | DESC ] ]
[ LIMIT { number | ALL } ]
[ OFFSET start ]
```

### Temas

- Cláusula WITH
- Lista SELECT
- EXCLUDE column_list
- Cláusula FROM
- Cláusula WHERE

- Cláusula GROUP BY
- Cláusula HAVING
- Cláusula QUALIFY
- UNION, INTERSECT y EXCEPT
- Cláusula ORDER BY
- Cláusula CONNECT BY
- Ejemplos de subconsultas
- Subconsultas correlacionadas

## Cláusula WITH

Una cláusula WITH es una cláusula opcional que precede a la lista SELECT en una consulta. La cláusula WITH define una o más common_table_expressions. Cada expresión común de tabla (CTE) define una tabla temporal, que es similar a la definición de una vista. Puede referenciar estas tablas temporales en la cláusula FROM. Solo se utilizan mientras se ejecuta la consulta a la que pertenecen. Cada CTE en la cláusula WITH especifica un nombre de tabla, una lista opcional de nombres de columnas y una expresión de consulta que toma el valor de una tabla (una instrucción SELECT). Cuando se referencia el nombre de la tabla temporal en la cláusula FROM de la misma expresión de consulta que lo define, la CTE es recursiva.

Las subconsultas de la cláusula WITH son una manera eficiente de definir tablas que puede utilizarse al ejecutar una única consulta. En todos los casos, se pueden obtener los mismos resultados al utilizar subconsultas en el cuerpo principal de la instrucción SELECT, pero las subconsultas de la cláusula WITH pueden resultar más sencillas de escribir y leer. Cuando es posible, las subconsultas de la cláusula WITH a las que se hace referencia varias veces se optimizan como subexpresiones comunes; es decir, puede ser posible evaluar una subconsulta WITH una vez y reutilizar sus resultados (tenga en cuenta que las subexpresiones comunes no se limitan a aquellas definidas en la cláusula WITH).

Sintaxis

```
[ WITH [RECURSIVE] common_table_expression [, common_table_expression , ...] ]
```

Donde common_table_expression puede ser recursiva o no recursiva. A continuación, se presenta la forma no recursiva:

```
CTE_table_name [ ( column_name [, ...] ) ] AS ( query )
```

A continuación, se presenta la forma recursiva de common_table_expression:

```
CTE_table_name (column_name [, ...] ) AS ( recursive_query )
```

Parámetros

RECURSIVE

Se trata de la palabra clave que identifica la consulta como una CTE recursiva. Esta palabra clave es necesaria si alguna common_table_expression definida en la cláusula WITH es recursiva. Solo puede especificar la palabra clave RECURSIVE una vez, inmediatamente después de la palabra clave WITH, incluso si la cláusula WITH contiene varias CTE recursivas. En general, una CTE recursiva es una subconsulta UNION ALL con dos partes.

common_table_expression

Define una tabla temporal a la que se puede referenciar en Cláusula FROM y se utiliza solo durante la ejecución de la consulta a la que pertenece.

CTE_table_name

Un nombre único para una tabla temporal que define los resultados de una subconsulta de la cláusula WITH. No se pueden usar nombres duplicados dentro de una cláusula WITH. Cada subconsulta debe tener un nombre de tabla al que se pueda hacer referencia en la Cláusula FROM.

column_name

Se trata de una lista de nombres de columnas de salida para la subconsulta de la cláusula WITH, separados por comas. La cantidad de nombres de columnas especificada debe ser igual o menor a la cantidad de columnas definidas por la subconsulta. Para una CTE que no es recursiva, column_name es opcional. Para una CTE recursiva, la lista de column_name es obligatoria.

consulta

Se trata de cualquier consulta SELECT compatible con Amazon Redshift. Consulte SELECT.

recursive_query

Una consulta UNION ALL que consta de dos subconsultas SELECT:

- La primera subconsulta SELECT no tiene una referencia recursiva al mismo CTE_table_name. Devuelve un conjunto de resultados que es la inicialización de la recursión. Esta parte se denomina miembro inicial o miembro de inicialización.
- La segunda subconsulta SELECT referencia el mismo CTE_table_name en su cláusula FROM. Esto se conoce como el miembro recursivo. La recursive_query contiene una condición WHERE para finalizar la recursive_query.

Notas de uso

Puede usar una cláusula WITH en las siguientes instrucciones SQL:

- SELECT
- SELECT INTO
- CREATE TABLE AS
- CREATE VIEW
- DECLARE
- EXPLAIN
- INSERT INTO...SELECT
- PREPARE
- UPDATE (dentro de una subconsulta de la cláusula WHERE. No puede definir una CTE recursiva en la subconsulta. La CTE recursiva debe ubicarse antes de la cláusula UPDATE).
- DELETE

Si la cláusula FROM de una consulta que contiene una cláusula WITH no referencia ninguna de las tablas definidas por la cláusula WITH, se ignora la cláusula WITH y la consulta se ejecuta como siempre.

Se puede hacer referencia a una tabla definida por una subconsulta de la cláusula WITH solo en el alcance de la consulta SELECT que inicia la cláusula WITH. Por ejemplo, se puede hacer referencia a dicha tabla en la cláusula FROM de una subconsulta en la lista SELECT, la cláusula WHERE o la cláusula HAVING. No se puede usar una cláusula WITH en una subconsulta y hacer referencia a su tabla en la cláusula FROM de una consulta principal o de otra subconsulta. Este patrón de consulta provoca un mensaje de error `relation table_name doesn't exist` para la tabla de la cláusula WITH.

No se puede especificar otra cláusula WITH dentro de una subconsulta de la cláusula WITH.

No se pueden realizar referencias futuras a tablas definidas por las subconsultas de la cláusula WITH. Por ejemplo, la siguiente consulta devuelve un error debido a la referencia futura a la tabla W2 en la definición de la tabla W1:

```
with w1 as (select * from w2), w2 as (select * from w1)
select * from sales;
ERROR:  relation "w2" does not exist
```

Una subconsulta de la cláusula WITH no puede constar de una instrucción SELECT INTO. No obstante, puede usar una cláusula WITH en una instrucción SELECT INTO.

Expresiones de tabla comunes recursivas

Una expresión común de tabla (CTE) recursiva es una CTE que se referencia a sí misma. Una CTE recursiva es útil para consultar datos jerárquicos, como gráficos de la organización que muestran las relaciones de informe entre los empleados y los administradores. Consulte Ejemplo: CTE recursiva.

Otro uso común es una factura de materiales multinivel, en la que un producto consta de muchos componentes, y cada componente, a su vez, consta de otros componentes o subconjuntos.

Asegúrese de limitar la profundidad de la recursividad al incluir una cláusula WHERE en la segunda subconsulta SELECT de la consulta recursiva. Para ver un ejemplo, consulta Ejemplo: CTE recursiva. De lo contrario, puede producirse un error similar a los siguientes:

- `Recursive CTE out of working buffers.`
- `Exceeded recursive CTE max rows limit, please add correct CTE termination predicates or change the max_recursion_rows parameter.`

> **Note**
>
> `max_recursion_rows` es un parámetro que establece el número máximo de filas que puede devolver una CTE recursiva para evitar bucles de recursión infinitos. No recomendamos cambiarlo por un valor superior al predeterminado. De este modo, se evita que los problemas de recursión infinita en las consultas ocupen un espacio excesivo en el clúster.

Puede especificar un criterio de ordenación y un límite en el resultado de la CTE recursiva. Puede incluir las opciones GROUP BY y DISTINCT en el resultado final de la CTE recursiva.

No se puede especificar una cláusula WITH RECURSIVE dentro de una subconsulta. El miembro de recursive_query no puede incluir una cláusula de ordenación o de límite.

Ejemplos

En el siguiente ejemplo, se muestra el caso posible más simple de una consulta que contiene una cláusula WITH. La consulta WITH denominada VENUECOPY selecciona todas las filas de la tabla VENUE. La consulta principal, a su vez, selecciona todas las filas de VENUECOPY. La tabla VENUECOPY existe solo durante esta consulta.

```
with venuecopy as (select * from venue)
select * from venuecopy order by 1 limit 10;
```

```
 venueid |         venuename          |    venuecity    | venuestate | venueseats
---------+----------------------------+-----------------+------------+------------
1 | Toyota Park               | Bridgeview      | IL         |          0
2 | Columbus Crew Stadium     | Columbus        | OH         |          0
3 | RFK Stadium               | Washington      | DC         |          0
4 | CommunityAmerica Ballpark | Kansas City     | KS         |          0
5 | Gillette Stadium          | Foxborough      | MA         |      68756
6 | New York Giants Stadium   | East Rutherford | NJ         |      80242
7 | BMO Field                 | Toronto         | ON         |          0
8 | The Home Depot Center     | Carson          | CA         |          0
9 | Dick's Sporting Goods Park | Commerce City   | CO         |          0
v     10 | Pizza Hut Park             | Frisco          | TX         |          0
(10 rows)
```

En el siguiente ejemplo, se muestra una cláusula WITH que produce dos tablas, denominadas VENUE_SALES y TOP_VENUES. La segunda tabla de la consulta WITH selecciona desde la primera. A su vez, la cláusula WHERE del bloque de la consulta principal contiene una subconsulta que limita la tabla TOP_VENUES.

```
with venue_sales as
(select venuename, venuecity, sum(pricepaid) as venuename_sales
from sales, venue, event
where venue.venueid=event.venueid and event.eventid=sales.eventid
group by venuename, venuecity),

top_venues as
(select venuename
from venue_sales
```

```
where venuename_sales > 800000)

select venuename, venuecity, venuestate,
sum(qtysold) as venue_qty,
sum(pricepaid) as venue_sales
from sales, venue, event
where venue.venueid=event.venueid and event.eventid=sales.eventid
and venuename in(select venuename from top_venues)
group by venuename, venuecity, venuestate
order by venuename;
```

```
        venuename       |   venuecity   | venuestate | venue_qty | venue_sales
------------------------+---------------+------------+-----------+-------------
August Wilson Theatre   | New York City | NY         |      3187 |  1032156.00
Biltmore Theatre        | New York City | NY         |      2629 |   828981.00
Charles Playhouse       | Boston        | MA         |      2502 |   857031.00
Ethel Barrymore Theatre | New York City | NY         |      2828 |   891172.00
Eugene O'Neill Theatre  | New York City | NY         |      2488 |   828950.00
Greek Theatre           | Los Angeles   | CA         |      2445 |   838918.00
Helen Hayes Theatre     | New York City | NY         |      2948 |   978765.00
Hilton Theatre          | New York City | NY         |      2999 |   885686.00
Imperial Theatre        | New York City | NY         |      2702 |   877993.00
Lunt-Fontanne Theatre   | New York City | NY         |      3326 |  1115182.00
Majestic Theatre        | New York City | NY         |      2549 |   894275.00
Nederlander Theatre     | New York City | NY         |      2934 |   936312.00
Pasadena Playhouse      | Pasadena      | CA         |      2739 |   820435.00
Winter Garden Theatre   | New York City | NY         |      2838 |   939257.00
(14 rows)
```

En los siguientes dos ejemplos se muestran las reglas para el alcance de las referencias de la tabla en función de las subconsultas de la cláusula WITH. La primera consulta se ejecuta, pero en la segunda se produce un error inesperado. La primera consulta tiene una subconsulta de la cláusula WITH dentro de la lista SELECT de la consulta principal. Se hace referencia a la tabla definida por la cláusula WITH (HOLIDAYS) en la cláusula FROM de la subconsulta de la lista SELECT:

```
select caldate, sum(pricepaid) as daysales,
(with holidays as (select * from date where holiday ='t')
select sum(pricepaid)
from sales join holidays on sales.dateid=holidays.dateid
where caldate='2008-12-25') as dec25sales
from sales join date on sales.dateid=date.dateid
where caldate in('2008-12-25','2008-12-31')
```

```
group by caldate
order by caldate;

caldate   | daysales | dec25sales
----------+----------+------------
2008-12-25 | 70402.00 |   70402.00
2008-12-31 | 12678.00 |   70402.00
(2 rows)
```

La segunda consulta falla porque intenta hacer referencia a la tabla HOLIDAYS en la consulta principal, así como en la subconsulta de la lista SELECT. Las referencias de la consulta principal están fuera de alcance.

```
select caldate, sum(pricepaid) as daysales,
(with holidays as (select * from date where holiday ='t')
select sum(pricepaid)
from sales join holidays on sales.dateid=holidays.dateid
where caldate='2008-12-25') as dec25sales
from sales join holidays on sales.dateid=holidays.dateid
where caldate in('2008-12-25','2008-12-31')
group by caldate
order by caldate;

ERROR:  relation "holidays" does not exist
```

Ejemplo: CTE recursiva

A continuación se muestra un ejemplo de CTE recursivo que devuelve los empleados que dependen directa o indirectamente de John. La consulta recursiva contiene una cláusula WHERE para limitar la profundidad de la recursividad a menos de 4 niveles.

```
--create and populate the sample table
  create table employee (
  id int,
  name varchar (20),
  manager_id int
  );

  insert into employee(id, name, manager_id)  values
(100, 'Carlos', null),
(101, 'John', 100),
(102, 'Jorge', 101),
```

```
(103, 'Kwaku', 101),
(110, 'Liu', 101),
(106, 'Mateo', 102),
(110, 'Nikki', 103),
(104, 'Paulo', 103),
(105, 'Richard', 103),
(120, 'Saanvi', 104),
(200, 'Shirley', 104),
(201, 'Sofía', 102),
(205, 'Zhang', 104);

--run the recursive query
  with recursive john_org(id, name, manager_id, level) as
( select id, name, manager_id, 1 as level
  from employee
  where name = 'John'
  union all
  select e.id, e.name, e.manager_id, level + 1 as next_level
  from employee e, john_org j
  where e.manager_id = j.id and level < 4
  )
 select distinct id, name, manager_id from john_org order by manager_id;
```

A continuación, se muestra el resultado de la consulta.

```
   id        name      manager_id
  ------+-----------+--------------
   101    John           100
   102    Jorge          101
   103    Kwaku          101
   110    Liu            101
   201    Sofía          102
   106    Mateo          102
   110    Nikki          103
   104    Paulo          103
   105    Richard        103
   120    Saanvi         104
   200    Shirley        104
   205    Zhang          104
```

A continuación, se muestra un gráfico de la organización del departamento de John.

# Lista SELECT

Temas


- Sintaxis
- Parámetros
- Notas de uso
- Ejemplos


La lista SELECT designa un nombre a las columnas, funciones y expresiones que desea que la consulta devuelva. La lista representa el resultado de la consulta.

Para obtener más información sobre las funciones SQL, consulte Referencia de funciones SQL. Para obtener más información sobre las expresiones, consulte Expresiones condicionales.

## Sintaxis

```
SELECT
[ TOP number ]
[ ALL | DISTINCT ] * | expression [ AS column_alias ] [, ...]
```

## Parámetros

TOP number

TOP toma un entero positivo como su argumento, lo que define la cantidad de filas que se devuelven al cliente. El comportamiento con la cláusula TOP es el mismo que el comportamiento con la cláusula LIMIT. El número de filas devueltas es fijo, pero el conjunto de filas no lo es. Para devolver un conjunto de filas coherente, utilice TOP o LIMIT junto con la cláusula ORDER BY.

ALL

Palabra clave redundante que define el comportamiento predeterminado si no se especifica DISTINCT. `SELECT ALL *` significa lo mismo que `SELECT *` (se seleccionen todas las filas de todas las columnas y se mantienen los duplicados).

DISTINCT

Opción que elimina las filas duplicadas del conjunto de resultados basándose en los valores coincidentes de una o más columnas.

> **Note**
>
> Si su aplicación admite claves externas o principales no válidas, algunas consultas podrían devolver resultados incorrectos. Por ejemplo, una consulta SELECT DISTINCT podría devolver filas duplicadas si la columna de clave principal no contiene todos los valores únicos. Para obtener más información acerca de estas restricciones, consulte Definición de restricciones de tabla.

* (asterisco)

Devuelve todo el contenido de la tabla (todas las columnas y filas).

expresión

Una expresión formada a partir de una o más columnas que existen en las tablas a las que hace referencia la consulta. Una expresión puede contener funciones SQL. Por ejemplo:

```
avg(datediff(day, listtime, saletime))
```

AS column_alias

Un nombre temporal para la columna que se utiliza en el conjunto de resultados finales. La palabra clave AS es opcional. Por ejemplo:

```
avg(datediff(day, listtime, saletime)) as avgwait
```

Si no se especifica un alias para una expresión que no sea un nombre de columna simple, el conjunto de resultados aplica un nombre predeterminado a esa columna.

> **Note**
>
> El alias se reconoce justo después de definirlo en la lista de destino. Puede usar un alias en otras expresiones definidas después de él en la misma lista objetivo. En el siguiente ejemplo, se ilustra este caso.
>
> ```
> select clicks / impressions as probability, round(100 * probability, 1) as
>  percentage from raw_data;
> ```

La referencia de alias lateral aporta un beneficio en el sentido de que no es necesario repetir la expresión con alias cuando se crean expresiones más complejas en la misma lista objetivo. Cuando Amazon Redshift analiza este tipo de referencia simplemente incorpora los alias definidos previamente. Si en la cláusula `FROM` hay una columna definida con el mismo nombre que en la expresión con alias anterior, tendrá prioridad la columna de la cláusula `FROM`. Por ejemplo en la consulta anterior si hay una columna denominada "probability" en la tabla raw_data, la "probability" de la segunda expresión en la lista objetivo hace referencia a dicha columna en vez de hacer referencia al nombre de alias "probability".

Notas de uso

TOP es una extensión SQL; proporciona una alternativa al comportamiento LIMIT. No se puede usar TOP y LIMIT en la misma consulta.

Ejemplos

En el siguiente ejemplo, se devuelven 10 filas desde la tabla SALES. Aunque la consulta usa la cláusula TOP, sigue devolviendo un conjunto de filas impredecible porque no se especifica ninguna cláusula ORDER BY,

```
select top 10 *
from sales;
```

La siguiente consulta es funcionalmente equivalente, pero usa una cláusula LIMIT en lugar de una cláusula TOP:

```
select *
from sales
limit 10;
```

En el siguiente ejemplo se devuelven las primeras 10 filas de la tabla SALES con la cláusula TOP, ordenadas por la columna QTYSOLD en orden descendente.

```
select top 10 qtysold, sellerid
from sales
order by qtysold desc, sellerid;
```

```
qtysold | sellerid
--------+----------
8 |      518
8 |      520
8 |      574
8 |      718
8 |      868
8 |     2663
8 |     3396
8 |     3726
8 |     5250
8 |     6216
(10 rows)
```

El ejemplo siguiente devuelve los primeros dos valores QTYSOLD y SELLERID de la tabla SALES, ordenados por la columna QTYSOLD:

```
select top 2 qtysold, sellerid
from sales
order by qtysold desc, sellerid;

qtysold | sellerid
--------+----------
8 |      518
8 |      520
(2 rows)
```

El siguiente ejemplo muestra la lista de grupos de categorías distintos de la tabla CATEGORY:

```
select distinct catgroup from category
order by 1;

catgroup
----------
Concerts
Shows
Sports
(3 rows)

--the same query, run without distinct
select catgroup from category
order by 1;
```

```
catgroup
----------
Concerts
Concerts
Concerts
Shows
Shows
Shows
Sports
Sports
Sports
Sports
Sports
(11 rows)
```

El ejemplo siguiente devuelve el conjunto distinto de números de semana de diciembre de 2008. Sin la cláusula DISTINCT, la instrucción devolvería 31 filas o una para cada día del mes.

```
select distinct week, month, year
from date
where month='DEC' and year=2008
order by 1, 2, 3;

week | month | year
-----+-------+------
49 | DEC   | 2008
50 | DEC   | 2008
51 | DEC   | 2008
52 | DEC   | 2008
53 | DEC   | 2008
(5 rows)
```

## EXCLUDE column_list

La cláusula EXCLUDE column_list asigna un nombre a las columnas que se excluyen de los resultados de la consulta. El uso de la opción EXCLUDE resulta útil cuando se necesita excluir solo un subconjunto de columnas de una tabla ancha, que es una tabla que contiene muchas columnas.

Temas

- Sintaxis

- Parámetros
- Ejemplos


### Sintaxis

```
EXCLUDE column_list
```

### Parámetros

column_list

Una lista separada por comas de uno o varios de los nombres de columna que existen en las tablas a las que hace referencia la consulta. De manera opcional, column_list se puede incluir entre paréntesis. En la lista de exclusiones de nombres de columna solo se admiten nombres de columna, no expresiones (como`upper(col1)`) o el asterisco (*).

```
column-name, ... | ( column-name, ... )
```

Por ejemplo:

```
SELECT * EXCLUDE col1, col2 FROM tablea;
```

```
SELECT * EXCLUDE (col1, col2) FROM tablea;
```

### Ejemplos

Los ejemplos siguientes utilizan la tabla SALES que contiene las columnas: salesid, listid, sellerid, buyerid, eventid, dateid, qtysold, pricepaid, commission y saletime. Para obtener más información sobre la tabla SALES, consulte Base de datos de muestra.

En el siguiente ejemplo, se devuelven filas de la tabla SALES, pero se excluye la columna SALETIME.

```
SELECT * EXCLUDE saletime FROM sales;

salesid | listid  | sellerid | buyerid | eventid | dateid  | qtysold  | pricepaid  |
 commission
```

```
--------+---------+----------+---------+---------+---------+----------+------------
+-----------
150314  | 173969  | 48680    | 816     | 8762    | 1827    | 2        | 688        |
 103.2
8325    | 8942    | 23600    | 1078    | 2557    | 1828    | 5        | 525        |
 78.75
46807   | 52711   | 34388    | 1047    | 2046    | 1828    | 2        | 482        |
 72.3
...
```

En el siguiente ejemplo, se devuelven filas de la tabla SALES, pero se excluyen las columnas QTYSOLD y SALETIME.

```
SELECT * EXCLUDE (qtysold, saletime) FROM sales;

salesid | listid  | sellerid | buyerid | eventid | dateid  | pricepaid  | commission
--------+---------+----------+---------+---------+---------+------------+-----------
150314  | 173969  | 48680    | 816     | 8762    | 1827    | 688        | 103.2
8325    | 8942    | 23600    | 1078    | 2557    | 1828    | 525        |  78.75
46807   | 52711   | 34388    | 1047    | 2046    | 1828    | 482        |  72.3
...
```

En el siguiente ejemplo, se crea una vista que devuelve filas de la tabla SALES, pero excluye la columna SALETIME.

```
CREATE VIEW sales_view AS SELECT * EXCLUDE saletime FROM sales;
SELECT * FROM sales_view;

salesid | listid  | sellerid | buyerid | eventid | dateid  | qtysold  | pricepaid  |
 commission
--------+---------+----------+---------+---------+---------+----------+------------
+-----------
150314  | 173969  | 48680    | 816     | 8762    | 1827    | 2        | 688        |
 103.2
8325    | 8942    | 23600    | 1078    | 2557    | 1828    | 5        | 525        |
 78.75
46807   | 52711   | 34388    | 1047    | 2046    | 1828    | 2        | 482        |
 72.3
...
```

En el siguiente ejemplo, se seleccionan únicamente las columnas que no están excluidas de una tabla temporal.

```
SELECT * EXCLUDE saletime INTO TEMP temp_sales FROM sales;
SELECT * FROM temp_sales;

salesid | listid  | sellerid | buyerid | eventid | dateid  | qtysold  | pricepaid  |
 commission
--------+---------+----------+---------+---------+---------+----------+------------
+-----------
150314  | 173969  | 48680    | 816     | 8762    | 1827    | 2        | 688        |
 103.2
8325    | 8942    | 23600    | 1078    | 2557    | 1828    | 5        | 525        |
 78.75
46807   | 52711   | 34388    | 1047    | 2046    | 1828    | 2        | 482        |
 72.3
...
```

## Cláusula FROM

La cláusula FROM en una consulta enumera las referencias de la tabla (tablas, vistas y subconsultas) desde las que se seleccionan los datos. Si se enumeran varias referencias de tabla, se deben combinar las tablas a través de la sintaxis adecuada en la cláusula FROM o en la cláusula WHERE. Si no se especifican criterios de combinación, el sistema procesa la consulta como una combinación cruzada (producto cartesiano).

Temas

- Sintaxis
- Parámetros
- Notas de uso
- Ejemplos de PIVOT y UNPIVOT
- Ejemplos de JOIN

### Sintaxis

```
FROM table_reference [, ...]
```

donde table_reference es uno de los siguientes:

```
with_subquery_table_name [ table_alias ]
table_name [ * ] [ table_alias ]
```

```
( subquery ) [ table_alias ]
table_reference [ NATURAL ] join_type table_reference
   [ ON join_condition | USING ( join_column [, ...] ) ]
table_reference PIVOT (
   aggregate(expr) [ [ AS ] aggregate_alias ]
   FOR column_name IN ( expression [ AS ] in_alias [, ...] )
) [ table_alias ]
table_reference UNPIVOT [ INCLUDE NULLS | EXCLUDE NULLS ] (
   value_column_name
   FOR name_column_name IN ( column_reference [ [ AS ]
   in_alias ] [, ...] )
) [ table_alias ]
UNPIVOT expression AS value_alias [ AT attribute_alias ]
```

El table_alias opcional se puede usar para dar nombres temporales a tablas y referencias de tablas complejas y, si se desea, también a sus columnas, como se muestra a continuación:

```
[ AS ] alias [ ( column_alias [, ...] ) ]
```

Parámetros

with_subquery_table_name

Una tabla definida por una subconsulta en la Cláusula WITH.

table_name

Nombre de una tabla o vista.

alias

Nombre alternativo temporal para una tabla o vista. Se debe proporcionar un alias para una tabla obtenida de una subconsulta. En otras referencias de tabla, los alias son opcionales La palabra clave AS es siempre opcional. Los alias de la tabla brindan un acceso directo para identificar tablas en otras partes de una consulta, como la cláusula WHERE. Por ejemplo:

```
select * from sales s, listing l
where s.listid=l.listid
```

column_alias

Nombre alternativo temporal para una columna en una tabla o vista.

subquery

Una expresión de consulta que toma el valor de una tabla. La tabla solo existe mientras dura la consulta y, por lo general, se le asigna un nombre o un alias. No obstante, no es obligatorio tener un alias. También puede definir nombres de columnas para tablas que derivan de subconsultas. Designar un nombre a los alias de las columnas es importante cuando desea combinar los resultados de las subconsultas con otras tablas y cuando desea seleccionar o limitar esas columnas en otros sitios de la consulta.

Una subconsulta puede contener una cláusula ORDER BY, pero es posible que esta cláusula no tenga ningún efecto si no se especifica también una cláusula OFFSET o LIMIT.

NATURAL

Define una combinación que utiliza automáticamente todos los pares de columnas con nombres idénticos en las dos tablas como las columnas de combinación. No se requiere una condición de combinación explícita. Por ejemplo, si las tablas CATEGORY y EVENT tienen columnas denominadas CATID, una combinación natural de estas tablas es una combinación de las columnas CATID.

> **Note**
>
> Si se especifica una combinación NATURAL, pero no existen pares de columnas con nombres idénticos en las tablas que deben combinarse, la consulta se establece en una combinación cruzada.

join_type

Especifique uno de los siguientes tipos de combinación:

- [INNER] JOIN
- LEFT [OUTER] JOIN
- RIGHT [OUTER] JOIN
- FULL [OUTER] JOIN
- CROSS JOIN

Las combinaciones cruzadas son combinaciones no calificadas; devuelven el producto cartesiano de dos tablas.

Las combinaciones internas y externas son combinaciones calificadas. Están calificadas implícitamente (en combinaciones naturales), con la sintaxis ON o USING en la cláusula FROM, o con una condición WHERE.

Una combinación interna devuelve filas coincidentes únicamente en función a la condición de combinación o a la lista de columnas de combinación. Una combinación externa devuelve todas las filas que la combinación interna equivalente devolvería, además de filas no coincidentes de la tabla "izquierda", tabla "derecha" o ambas tablas. La tabla izquierda es la primera tabla de la lista, y la tabla derecha es la segunda tabla de la lista. Las filas no coincidentes contienen valores NULL para llenar el vacío de las columnas de salida.

ON join_condition

Especificación del tipo de combinación donde las columnas de combinación se establecen como una condición que sigue la palabra clave ON. Por ejemplo:

```
sales join listing
on sales.listid=listing.listid and sales.eventid=listing.eventid
```

USING ( join_column [, ...] )

Especificación del tipo de combinación donde las columnas de combinación aparecen enumeradas entre paréntesis. Si se especifican varias columnas de combinación, se delimitan por comas. La palabra clave USING debe preceder a la lista. Por ejemplo:

```
sales join listing
using (listid,eventid)
```

PIVOT

Rota la salida de filas a columnas, con el fin de representar datos tabulares en un formato fácil de leer. La salida se representa horizontalmente en varias columnas. PIVOT es similar a una consulta GROUP BY con una agregación, utilizando una expresión de agregación para especificar un formato de salida. No obstante, a diferencia de GROUP BY, los resultados se devuelven en columnas en lugar de filas.

Para ver ejemplos que muestran cómo realizar consultas con PIVOT y UNPIVOT, consulte Ejemplos de PIVOT y UNPIVOT.

UNPIVOT

Rotación de columnas en filas con UNPIVOT: el operador transforma las columnas de resultados, de una tabla de entrada o resultados de consultas, en filas, para facilitar la lectura de la salida. UNPIVOT combina los datos de sus columnas de entrada en dos columnas de resultados: una columna de nombres y una columna de valores. La columna de nombres contiene nombres de columna de la entrada, como entradas de fila. La columna de valores contiene valores de las columnas de entrada, como los resultados de una agregación. Por ejemplo, el recuento de elementos de varias categorías.

Anulación de la dinamización de objetos con UNPIVOT (SUPER): puede realizar una anulación de la dinamización de objetos, donde la expresión es una expresión SUPER que hace referencia a otro elemento de la cláusula FROM. Para obtener más información, consulte Anulación de la dinamización de los objetos. Tiene también ejemplos que muestran cómo realizar consultas de datos semiestructurados, como datos con formato JSON.

Notas de uso

Las columnas de combinación deben tener tipos de datos comparables.

Una combinación NATURAL o USING retiene solo uno de cada par de columnas de combinación en el conjunto de resultados intermedios.

Una combinación con la sintaxis ON retiene ambas columnas de combinación en su conjunto de resultados intermedios.

Véase también Cláusula WITH.

Ejemplos de PIVOT y UNPIVOT

PIVOT y UNPIVOT son parámetros de la cláusula FROM que alternan la salida de la consulta de filas a columnas y de columnas a filas, respectivamente. Representan resultados de consultas tabulares en un formato fácil de leer. En los siguientes ejemplos se utilizan datos y consultas de prueba para mostrar cómo utilizarlos.

Para obtener más información sobre estos parámetros y sus opciones predeterminadas, consulte Cláusula FROM.

### Ejemplos de PIVOT

Configure la tabla y los datos de muestra y utilícelos para ejecutar las consultas de ejemplo posteriores.

```
CREATE TABLE part (
    partname varchar,
    manufacturer varchar,
    quality int,
    price decimal(12, 2)
);

INSERT INTO part VALUES ('prop', 'local parts co', 2, 10.00);
INSERT INTO part VALUES ('prop', 'big parts co', NULL, 9.00);
INSERT INTO part VALUES ('prop', 'small parts co', 1, 12.00);

INSERT INTO part VALUES ('rudder', 'local parts co', 1, 2.50);
INSERT INTO part VALUES ('rudder', 'big parts co', 2, 3.75);
INSERT INTO part VALUES ('rudder', 'small parts co', NULL, 1.90);

INSERT INTO part VALUES ('wing', 'local parts co', NULL, 7.50);
INSERT INTO part VALUES ('wing', 'big parts co', 1, 15.20);
INSERT INTO part VALUES ('wing', 'small parts co', NULL, 11.80);
```

PIVOT en `partname` con una agregación `AVG` en `price`.

```
SELECT *
FROM (SELECT partname, price FROM part) PIVOT (
    AVG(price) FOR partname IN ('prop', 'rudder', 'wing')
);
```

La consulta genera el siguiente resultado.

```
  prop   |  rudder  |  wing
---------+----------+---------
 10.33   | 2.71     | 11.50
```

En el ejemplo anterior, los resultados se transforman en columnas. En el siguiente ejemplo se muestra una consulta de `GROUP BY` que devuelve los precios promedio en filas, en lugar de hacerlo en columnas.

```
SELECT partname, avg(price)
```

```
FROM (SELECT partname, price FROM part)
WHERE partname IN ('prop', 'rudder', 'wing')
GROUP BY partname;
```

La consulta genera el siguiente resultado.

```
 partname |  avg
----------+-------
 prop     | 10.33
 rudder   |  2.71
 wing     | 11.50
```

Un ejemplo de PIVOT con `manufacturer` como columna implícita.

```
SELECT *
FROM (SELECT quality, manufacturer FROM part) PIVOT (
    count(*) FOR quality IN (1, 2, NULL)
);
```

La consulta genera el siguiente resultado.

```
 manufacturer      | 1  | 2  | null
-------------------+----+----+------
 local parts co    | 1  | 1  |  1
 big parts co      | 1  | 1  |  1
 small parts co    | 1  | 0  |  2
```

Columnas de tabla de entrada a las que no se hace referencia en la definición de PIVOT se agregan implícitamente a la tabla de resultados. Este es el caso de la columna `manufacturer` del ejemplo anterior. En el ejemplo también se muestra que NULL es un valor válido para el operador IN.

PIVOT en el ejemplo anterior devuelve información similar a la siguiente consulta, lo que incluye GROUP BY. La diferencia es que PIVOT devuelve el valor `0` para la columna 2 y el fabricante `small parts co`. La consulta GROUP BY no contiene una fila correspondiente. En la mayoría de los casos, PIVOT inserta NULL si una fila no tiene datos de entrada para una columna determinada. No obstante, el agregado de recuento no devuelve NULL y `0` es el valor predeterminado.

```
SELECT manufacturer, quality, count(*)
FROM (SELECT quality, manufacturer FROM part)
WHERE quality IN (1, 2) OR quality IS NULL
GROUP BY manufacturer, quality
```

```
ORDER BY manufacturer;
```

La consulta genera el siguiente resultado.

```
 manufacturer        | quality | count
---------------------+---------+-------
 big parts co        |         |     1
 big parts co        |       2 |     1
 big parts co        |       1 |     1
 local parts co      |       2 |     1
 local parts co      |       1 |     1
 local parts co      |         |     1
 small parts co      |       1 |     1
 small parts co      |         |     2
```

El operador PIVOT acepta alias opcionales en la expresión agregada y en cada valor para el operador IN. Utilice alias para personalizar los nombres de las columnas. Si no hay alias agregados, solo se utilizan alias de la lista IN. De lo contrario, el alias agregado se agrega al nombre de la columna con un guion bajo para separar los nombres.

```
SELECT *
FROM (SELECT quality, manufacturer FROM part) PIVOT (
    count(*) AS count FOR quality IN (1 AS high, 2 AS low, NULL AS na)
);
```

La consulta genera el siguiente resultado.

```
 manufacturer      | high_count  | low_count | na_count
-------------------+-------------+-----------+----------
 local parts co    |           1 |         1 |        1
 big parts co      |           1 |         1 |        1
 small parts co    |           1 |         0 |        2
```

Configure la tabla y los datos de muestra siguientes y utilícelos para ejecutar las consultas de ejemplo posteriores. Los datos representan las fechas de reserva de una colección de hoteles.

```
CREATE TABLE bookings (
    booking_id int,
    hotel_code char(8),
    booking_date date,
    price decimal(12, 2)
```

```
);

INSERT INTO bookings VALUES (1, 'FOREST_L', '02/01/2023', 75.12);
INSERT INTO bookings VALUES (2, 'FOREST_L', '02/02/2023', 75.00);
INSERT INTO bookings VALUES (3, 'FOREST_L', '02/04/2023', 85.54);

INSERT INTO bookings VALUES (4, 'FOREST_L', '02/08/2023', 75.00);
INSERT INTO bookings VALUES (5, 'FOREST_L', '02/11/2023', 75.00);
INSERT INTO bookings VALUES (6, 'FOREST_L', '02/14/2023', 90.00);

INSERT INTO bookings VALUES (7, 'FOREST_L', '02/21/2023', 60.00);
INSERT INTO bookings VALUES (8, 'FOREST_L', '02/22/2023', 85.00);
INSERT INTO bookings VALUES (9, 'FOREST_L', '02/27/2023', 90.00);

INSERT INTO bookings VALUES (10, 'DESERT_S', '02/01/2023', 98.00);
INSERT INTO bookings VALUES (11, 'DESERT_S', '02/02/2023', 75.00);
INSERT INTO bookings VALUES (12, 'DESERT_S', '02/04/2023', 85.00);

INSERT INTO bookings VALUES (13, 'DESERT_S', '02/05/2023', 75.00);
INSERT INTO bookings VALUES (14, 'DESERT_S', '02/06/2023', 34.00);
INSERT INTO bookings VALUES (15, 'DESERT_S', '02/09/2023', 85.00);

INSERT INTO bookings VALUES (16, 'DESERT_S', '02/12/2023', 23.00);
INSERT INTO bookings VALUES (17, 'DESERT_S', '02/13/2023', 76.00);
INSERT INTO bookings VALUES (18, 'DESERT_S', '02/14/2023', 85.00);

INSERT INTO bookings VALUES (19, 'OCEAN_WV', '02/01/2023', 98.00);
INSERT INTO bookings VALUES (20, 'OCEAN_WV', '02/02/2023', 75.00);
INSERT INTO bookings VALUES (21, 'OCEAN_WV', '02/04/2023', 85.00);

INSERT INTO bookings VALUES (22, 'OCEAN_WV', '02/06/2023', 75.00);
INSERT INTO bookings VALUES (23, 'OCEAN_WV', '02/09/2023', 34.00);
INSERT INTO bookings VALUES (24, 'OCEAN_WV', '02/12/2023', 85.00);

INSERT INTO bookings VALUES (25, 'OCEAN_WV', '02/13/2023', 23.00);
INSERT INTO bookings VALUES (26, 'OCEAN_WV', '02/14/2023', 76.00);
INSERT INTO bookings VALUES (27, 'OCEAN_WV', '02/16/2023', 85.00);

INSERT INTO bookings VALUES (28, 'CITY_BLD', '02/01/2023', 98.00);
INSERT INTO bookings VALUES (29, 'CITY_BLD', '02/02/2023', 75.00);
INSERT INTO bookings VALUES (30, 'CITY_BLD', '02/04/2023', 85.00);

INSERT INTO bookings VALUES (31, 'CITY_BLD', '02/12/2023', 75.00);
INSERT INTO bookings VALUES (32, 'CITY_BLD', '02/13/2023', 34.00);
```

```
INSERT INTO bookings VALUES (33, 'CITY_BLD', '02/17/2023', 85.00);

INSERT INTO bookings VALUES (34, 'CITY_BLD', '02/22/2023', 23.00);
INSERT INTO bookings VALUES (35, 'CITY_BLD', '02/23/2023', 76.00);
INSERT INTO bookings VALUES (36, 'CITY_BLD', '02/24/2023', 85.00);
```

En este ejemplo de consulta, los registros de reserva se suman para obtener un total de cada semana. La fecha de finalización de cada semana se convierte en un nombre de columna.

```
SELECT * FROM
    (SELECT
       booking_id,
       (date_trunc('week', booking_date::date) + '5 days'::interval)::date as enddate,
       hotel_code AS "hotel code"
FROM bookings
) PIVOT (
    count(booking_id) FOR enddate IN ('2023-02-04','2023-02-11','2023-02-18')
);
```

La consulta genera el siguiente resultado.

```
 hotel code | 2023-02-04  | 2023-02-11 | 2023-02-18
------------+-------------+------------+----------
 FOREST_L   |           3 |          2 |        1
 DESERT_S   |           4 |          3 |        2
 OCEAN_WV   |           3 |          3 |        3
 CITY_BLD   |           3 |          1 |        2
```

Amazon Redshift no admite CROSSTAB para dinamizar en varias columnas. Pero puede cambiar los datos de filas a columnas, de forma similar a una agregación con PIVOT, con una consulta como la siguiente. Para ello se utilizan los mismos datos de muestra de reservas que en el ejemplo anterior.

```
SELECT
  booking_date,
  MAX(CASE WHEN hotel_code = 'FOREST_L' THEN 'forest is booked' ELSE '' END) AS
 FOREST_L,
  MAX(CASE WHEN hotel_code = 'DESERT_S' THEN 'desert is booked' ELSE '' END) AS
 DESERT_S,
  MAX(CASE WHEN hotel_code = 'OCEAN_WV' THEN 'ocean is booked' ELSE '' END)  AS
 OCEAN_WV
FROM bookings
GROUP BY booking_date
```

```
ORDER BY booking_date asc;
```

La consulta de ejemplo da como resultado las fechas de reserva que aparecen junto a frases cortas que indican qué hoteles están reservados.

```
 booking_date  | forest_l         | desert_s         | ocean_wv
---------------+------------------+------------------+--------------------
 2023-02-01    | forest is booked | desert is booked |  ocean is booked
 2023-02-02    | forest is booked | desert is booked |  ocean is booked
 2023-02-04    | forest is booked | desert is booked |  ocean is booked
 2023-02-05    |                  | desert is booked |
 2023-02-06    |                  | desert is booked |
```

Las siguientes notas de uso se indican a continuación para PIVOT:

- PIVOT se puede aplicar a tablas, subconsultas y expresiones de tabla comunes (CTE). PIVOT no se puede aplicar a ninguna expresión JOIN, CTE recursivos, expresiones PIVOT o UNPIVOT. Tampoco se admiten expresiones SUPER sin anidar ni tablas anidadas Redshift Spectrum.
- PIVOT admite las funciones de agregación COUNT, SUM, MIN, MAX y AVG.
- La expresión de agregación PIVOT tiene que ser una llamada de una función de agregación admitida. No se admiten expresiones complejas además de las de agregación. Los argumentos de agregación no pueden contener referencias a tablas distintas de las tablas PIVOT de entrada. Tampoco se admiten las referencias correlacionadas a una consulta principal. El argumento agregado puede contener subconsultas. Se pueden correlacionar internamente o en la tabla de entrada PIVOT.
- Los valores de la lista PIVOT IN no pueden ser referencias de columna ni subconsultas. Cada valor debe ser compatible con el tipo de referencia de columna FOR.
- Si los valores de la lista IN no tienen alias, PIVOT genera nombres de columna predeterminados. Para valores IN constantes como “abc” o 5, el nombre de columna predeterminado es la constante en sí. Para cualquier expresión compleja, el nombre de la columna es un nombre predeterminado estándar de Amazon Redshift, como ?column?.

### Ejemplos de UNPIVOT

Configure los datos de muestra y utilícelos para ejecutar los ejemplos posteriores.

```
CREATE TABLE count_by_color (quality varchar, red int, green int, blue int);
```

```
INSERT INTO count_by_color VALUES ('high', 15, 20, 7);
INSERT INTO count_by_color VALUES ('normal', 35, NULL, 40);
INSERT INTO count_by_color VALUES ('low', 10, 23, NULL);
```

UNPIVOT en las columnas de entrada roja, verde y azul.

```
SELECT *
FROM (SELECT red, green, blue FROM count_by_color) UNPIVOT (
    cnt FOR color IN (red, green, blue)
);
```

La consulta genera el siguiente resultado.

```
 color | cnt
-------+-----
 red   |  15
 red   |  35
 red   |  10
 green |  20
 green |  23
 blue  |   7
 blue  |  40
```

De manera predeterminada, los valores NULL de la columna de entrada se omiten y no producen una fila de resultados.

En el siguiente ejemplo, se muestra UNPIVOT con INCLUDE NULLS.

```
SELECT *
FROM (
    SELECT red, green, blue
    FROM count_by_color
) UNPIVOT INCLUDE NULLS (
    cnt FOR color IN (red, green, blue)
);
```

Se genera el siguiente resultado.

```
 color | cnt
-------+-----
 red   |  15
```

```
 red   |  35
 red   |  10
 green |  20
 green |
 green |  23
 blue  |   7
 blue  |  40
 blue  |
```

Si se configura el parámetro `INCLUDING NULLS`, los valores de entrada `NULL` generan filas de resultados.

Un ejemplo de `The following query shows UNPIVOT` con `quality` como columna implícita.

```
SELECT *
FROM count_by_color UNPIVOT (
    cnt FOR color IN (red, green, blue)
);
```

La consulta genera el siguiente resultado.

```
 quality | color | cnt
---------+-------+-----
 high    | red   |  15
 normal  | red   |  35
 low     | red   |  10
 high    | green |  20
 low     | green |  23
 high    | blue  |   7
 normal  | blue  |  40
```

Columnas de tabla de entrada a las que no se hace referencia en la definición de `UNPIVOT`, se agregan implícitamente a la tabla de resultados. Este es el caso de la columna `quality` del ejemplo.

En el siguiente ejemplo, se muestra `UNPIVOT` con alias para valores en la lista `IN`.

```
SELECT *
FROM count_by_color UNPIVOT (
    cnt FOR color IN (red AS r, green AS g, blue AS b)
);
```

La consulta anterior genera el siguiente resultado.

```
 quality | color | cnt
---------+-------+-----
 high    | r     |  15
 normal  | r     |  35
 low     | r     |  10
 high    | g     |  20
 low     | g     |  23
 high    | b     |   7
 normal  | b     |  40
```

El operador UNPIVOT acepta alias opcionales en cada valor de la lista IN. Cada alias proporciona personalización de los datos de cada columna value.

Las siguientes notas de uso se indican a continuación para UNPIVOT.

- UNPIVOT se puede aplicar a tablas, subconsultas y expresiones de tabla comunes (CTE). UNPIVOT no se puede aplicar a ninguna expresión JOIN, CTE recursivos, expresiones PIVOT o UNPIVOT. Tampoco se admiten expresiones SUPER sin anidar ni tablas anidadas Redshift Spectrum.
- La lista UNPIVOT IN debe contener solo referencias de columna de tabla de entrada. Las columnas de la lista IN deben tener un tipo común con el que todas sean compatibles. La columna de valor UNPIVOT tiene este tipo común. La columna de nombre UNPIVOT es de tipo VARCHAR.
- Si un valor de la lista IN no tiene un alias, UNPIVOT utiliza el nombre de la columna como valor predeterminado.

### Ejemplos de JOIN

Se utiliza una cláusula JOIN de SQL para combinar los datos de dos o más tablas en función de los campos comunes. Es posible que los resultados cambien o no cambien según el método de combinación especificado. Para obtener más información acerca de la sintaxis de la cláusula JOIN, consulte Parámetros.

En los siguientes ejemplos se usan datos de los ejemplos de TICKIT. Para obtener más información acerca del esquema de base de datos, consulte Base de datos de muestra. Para obtener información sobre cómo cargar datos de ejemplo, consulte Carga de datos en la Guía de introducción a Amazon Redshift.

La siguiente consulta es una combinación interna (sin la palabra clave JOIN) entre la tabla LISTING y la tabla SALES, donde LISTID de la tabla LISTING está entre 1 y 5. Esta consulta relaciona los

valores de la columna LISTID en la tabla LISTING (la tabla izquierda) y la tabla SALES (la tabla derecha). Los resultados muestran que LISTID 1, 4 y 5 coinciden con los criterios.

```
select listing.listid, sum(pricepaid) as price, sum(commission) as comm
from listing, sales
where listing.listid = sales.listid
and listing.listid between 1 and 5
group by 1
order by 1;

listid | price  |  comm
-------+--------+--------
     1 | 728.00 | 109.20
     4 |  76.00 |  11.40
     5 | 525.00 |  78.75
```

La siguiente consulta es una combinación externa izquierda. Las combinaciones externas izquierdas y derechas conservan valores de una de las tablas combinadas cuando no se encuentra una coincidencia en la otra tabla. Las tablas izquierda y derecha son la primera tabla y la segunda tabla que aparecen en la sintaxis. Los valores NULL se utilizan para rellenar los "espacios" en el conjunto de resultados. Esta consulta relaciona los valores de la columna LISTID en la tabla LISTING (la tabla izquierda) y la tabla SALES (la tabla derecha). Los resultados muestran que LISTID 2 y 3 no tienen ventas.

```
select listing.listid, sum(pricepaid) as price, sum(commission) as comm
from listing left outer join sales on sales.listid = listing.listid
where listing.listid between 1 and 5
group by 1
order by 1;

listid | price  |  comm
-------+--------+--------
     1 | 728.00 | 109.20
     2 | NULL   | NULL
     3 | NULL   | NULL
     4 |  76.00 |  11.40
     5 | 525.00 |  78.75
```

La siguiente consulta es una combinación externa derecha. Esta consulta relaciona los valores de la columna LISTID en la tabla LISTING (la tabla izquierda) y la tabla SALES (la tabla derecha). Los resultados muestran que LISTID 1, 4 y 5 coinciden con los criterios.

```
select listing.listid, sum(pricepaid) as price, sum(commission) as comm
from listing right outer join sales on sales.listid = listing.listid
where listing.listid between 1 and 5
group by 1
order by 1;

listid | price  |  comm
-------+--------+--------
     1 | 728.00 | 109.20
     4 |  76.00 |  11.40
     5 | 525.00 |  78.75
```

La siguiente consulta es una combinación completa. Las combinaciones completas retienen valores de las tablas combinadas cuando no se encuentra una coincidencia en la otra tabla. Las tablas izquierda y derecha son la primera tabla y la segunda tabla que aparecen en la sintaxis. Los valores NULL se utilizan para rellenar los "espacios" en el conjunto de resultados. Esta consulta relaciona los valores de la columna LISTID en la tabla LISTING (la tabla izquierda) y la tabla SALES (la tabla derecha). Los resultados muestran que LISTID 2 y 3 no tienen ventas.

```
select listing.listid, sum(pricepaid) as price, sum(commission) as comm
from listing full join sales on sales.listid = listing.listid
where listing.listid between 1 and 5
group by 1
order by 1;

listid | price  |  comm
-------+--------+--------
     1 | 728.00 | 109.20
     2 | NULL   | NULL
     3 | NULL   | NULL
     4 |  76.00 |  11.40
     5 | 525.00 |  78.75
```

La siguiente consulta es una combinación completa. Esta consulta relaciona los valores de la columna LISTID en la tabla LISTING (la tabla izquierda) y la tabla SALES (la tabla derecha). Solo se encuentran en los resultados filas que no dan lugar a ninguna venta (LISTID 2 y 3).

```
select listing.listid, sum(pricepaid) as price, sum(commission) as comm
from listing full join sales on sales.listid = listing.listid
where listing.listid between 1 and 5
and (listing.listid IS NULL or sales.listid IS NULL)
```

```
group by 1
order by 1;

listid | price  |  comm
-------+--------+--------
     2 | NULL   | NULL
     3 | NULL   | NULL
```

El siguiente ejemplo es una combinación interna con la cláusula ON. En este caso, las filas NULL no se devuelven.

```
select listing.listid, sum(pricepaid) as price, sum(commission) as comm
from sales join listing
on sales.listid=listing.listid and sales.eventid=listing.eventid
where listing.listid between 1 and 5
group by 1
order by 1;

listid | price  |  comm
-------+--------+--------
     1 | 728.00 | 109.20
     4 |  76.00 |  11.40
     5 | 525.00 |  78.75
```

La siguiente consulta es una combinación cruzada o cartesiana de la tabla LISTING y la tabla SALES con un predicado para limitar los resultados. Esta consulta coincide con los valores de columna LISTID de la tabla SALES y la tabla LISTING para LISTID 1, 2, 3, 4 y 5 de ambas tablas. Los resultados muestran que 20 filas coinciden con los criterios.

```
select sales.listid as sales_listid, listing.listid as listing_listid
from sales cross join listing
where sales.listid between 1 and 5
and listing.listid between 1 and 5
order by 1,2;

sales_listid | listing_listid
-------------+---------------
1            | 1
1            | 2
1            | 3
1            | 4
1            | 5
```

```
4            | 1
4            | 2
4            | 3
4            | 4
4            | 5
5            | 1
5            | 1
5            | 2
5            | 2
5            | 3
5            | 3
5            | 4
5            | 4
5            | 5
5            | 5
```

El ejemplo siguiente es una combinación natural entre dos tablas. En este caso, las columnas listid, sellerid, eventid y dateid tienen nombres y tipos de datos idénticos en ambas tablas y, por lo tanto, se utilizan como columnas de combinación. Los resultados tienen un límite de cinco filas.

```
select listid, sellerid, eventid, dateid, numtickets
from listing natural join sales
order by 1
limit 5;

listid | sellerid  | eventid | dateid | numtickets
-------+-----------+---------+--------+-----------
113    | 29704     | 4699    | 2075   | 22
115    | 39115     | 3513    | 2062   | 14
116    | 43314     | 8675    | 1910   | 28
118    | 6079      | 1611    | 1862   | 9
163    | 24880     | 8253    | 1888   | 14
```

El ejemplo siguiente es una combinación entre dos tablas con la cláusula USING. En este caso, las columnas listid y eventid se utilizan como columnas de combinación. Los resultados tienen un límite de cinco filas.

```
select listid, listing.sellerid, eventid, listing.dateid, numtickets
from listing join sales
using (listid, eventid)
order by 1
limit 5;
```

```
listid | sellerid | eventid | dateid | numtickets
-------+----------+---------+--------+-----------
1      | 36861    | 7872    | 1850   | 10
4      | 8117     | 4337    | 1970   | 8
5      | 1616     | 8647    | 1963   | 4
5      | 1616     | 8647    | 1963   | 4
6      | 47402    | 8240    | 2053   | 18
```

La siguiente consulta es una combinación interna de dos subconsultas en la cláusula FROM. La consulta busca la cantidad de tickets vendidos y sin vender para diferentes categorías de eventos (conciertos y espectáculos). Estas subconsultas de la cláusula FROM son subconsultas de tabla; pueden devolver varias columnas y filas.

```
select catgroup1, sold, unsold
from
(select catgroup, sum(qtysold) as sold
from category c, event e, sales s
where c.catid = e.catid and e.eventid = s.eventid
group by catgroup) as a(catgroup1, sold)
join
(select catgroup, sum(numtickets)-sum(qtysold) as unsold
from category c, event e, sales s, listing l
where c.catid = e.catid and e.eventid = s.eventid
and s.listid = l.listid
group by catgroup) as b(catgroup2, unsold)

on a.catgroup1 = b.catgroup2
order by 1;

catgroup1 |  sold  | unsold
----------+--------+--------
Concerts  | 195444 |1067199
Shows     | 149905 | 817736
```

## Cláusula WHERE

La cláusula WHERE contiene condiciones que combinan tablas o que aplican predicados a columnas de las tablas. Las tablas pueden combinarse de manera interna a través de la sintaxis adecuada en la cláusula WHERE o en la cláusula FROM. Los criterios de combinación externa deben especificarse en la cláusula FROM.

Sintaxis

```
[ WHERE condition ]
```

condition

Cualquier condición de búsqueda con un resultado booleano, como una condición de combinación o un predicado en una columna de la tabla. Los siguientes ejemplos son condiciones de combinación válidas:

```
sales.listid=listing.listid
sales.listid<>listing.listid
```

Los siguientes ejemplos son condiciones válidas de columnas en tablas:

```
catgroup like 'S%'
venueseats between 20000 and 50000
eventname in('Jersey Boys','Spamalot')
year=2008
length(catdesc)>25
date_part(month, caldate)=6
```

Las condiciones pueden ser simples o complejas. Para las condiciones complejas, puede utilizar paréntesis para aislar las unidades lógicas. En el siguiente ejemplo, la condición de combinación está entre paréntesis.

```
where (category.catid=event.catid) and category.catid in(6,7,8)
```

Notas de uso

Puede usar alias en la cláusula WHERE para hacer referencia a expresiones de listas de selección.

No puede limitar los resultados de las funciones de agregación en la cláusula WHERE; utilice la cláusula HAVING con este fin.

Las columnas que están limitadas en la cláusula WHERE deben derivar de referencias de tabla en la cláusula FROM.

Ejemplo

La siguiente consulta utiliza una combinación de diferentes restricciones de la cláusula WHERE, incluida una condición de combinación para las tablas SALES y EVENT, un predicado en la columna EVENTNAME y dos predicados en la columna STARTTIME.

```
select eventname, starttime, pricepaid/qtysold as costperticket, qtysold
from sales, event
where sales.eventid = event.eventid
and eventname='Hannah Montana'
and date_part(quarter, starttime) in(1,2)
and date_part(year, starttime) = 2008
order by 3 desc, 4, 2, 1 limit 10;

eventname    |      starttime      |   costperticket   | qtysold
---------------+---------------------+-------------------+---------
Hannah Montana | 2008-06-07 14:00:00 |     1706.00000000 |       2
Hannah Montana | 2008-05-01 19:00:00 |     1658.00000000 |       2
Hannah Montana | 2008-06-07 14:00:00 |     1479.00000000 |       1
Hannah Montana | 2008-06-07 14:00:00 |     1479.00000000 |       3
Hannah Montana | 2008-06-07 14:00:00 |     1163.00000000 |       1
Hannah Montana | 2008-06-07 14:00:00 |     1163.00000000 |       2
Hannah Montana | 2008-06-07 14:00:00 |     1163.00000000 |       4
Hannah Montana | 2008-05-01 19:00:00 |      497.00000000 |       1
Hannah Montana | 2008-05-01 19:00:00 |      497.00000000 |       2
Hannah Montana | 2008-05-01 19:00:00 |      497.00000000 |       4
(10 rows)
```

Combinaciones externas de estilo Oracle en la cláusula WHERE

Para la compatibilidad con Oracle, Amazon Redshift admite el operador de combinación externa de Oracle (+) en condiciones de combinación de la cláusula WHERE. Este operador está diseñado para utilizarse exclusivamente en la definición de condiciones de combinación externa; no intente utilizarlo en otros contextos. En la mayoría de casos, otros usos de este operador se omiten silenciosamente.

Una combinación externa devuelve todas las filas que la combinación interna equivalente devolvería, además de filas no coincidentes de una o ambas tablas. En la cláusula FROM, puede especificar combinaciones externas izquierdas, derechas y completas. En la cláusula WHERE, puede especificar combinaciones externas izquierdas y derechas únicamente.

Para realizar una combinación externa en las tablas TABLE1 y TABLE2, y devolver filas no coincidentes de la TABLE1 (una combinación externa izquierda), especifique `TABLE1 LEFT`

`OUTER JOIN TABLE2` en la cláusula FROM o aplique el operador (+) en todas las columnas de combinación de la TABLE2 en la cláusula WHERE. Para todas las filas en la TABLE1 que tengan filas no coincidentes en la TABLE2, el resultado de la consulta contiene valores nulos para cualquier expresión de lista de selección que contiene columnas de la TABLE2.

Para producir el mismo comportamiento para todas las filas en la TABLE2 que no tienen filas coincidentes en la TABLE1, especifique `TABLE1 RIGHT OUTER JOIN TABLE2` en la cláusula FROM o aplique el operador (+) en todas las columnas de combinación de la TABLE1 en la cláusula WHERE.

Sintaxis básica

```
[ WHERE {
[ table1.column1 = table2.column1(+) ]
[ table1.column1(+) = table2.column1 ]
}
```

La primera condición equivale a lo siguiente:

```
from table1 left outer join table2
on table1.column1=table2.column1
```

La segunda condición equivale a lo siguiente:

```
from table1 right outer join table2
on table1.column1=table2.column1
```

> **Note**
>
> La sintaxis que se muestra aquí cubre el caso simple de una equijoin sobre un par de columnas de combinación. No obstante, también son válidos otros tipos de condiciones de comparación y varios pares de columnas de combinación.

Por ejemplo, la siguiente cláusula WHERE define una combinación externa sobre dos pares de columnas. El operador (+) debe estar adjunto a la misma tabla en ambas condiciones:

```
where table1.col1 > table2.col1(+)
```

```
and table1.col2 = table2.col2(+)
```

Notas de uso

Siempre que sea posible, use la sintaxis OUTER JOIN de la cláusula FROM estándar en lugar del operador (+) en la cláusula WHERE. Las consultas que contienen el operador (+) están sujetas a las siguientes reglas:

- Solo se puede usar el operador (+) en la cláusula WHERE, y solo en referencia a las columnas de tablas o vistas.
- No se puede aplicar el operador (+) a las expresiones. No obstante, una expresión puede contener columnas que utilicen el operador (+). Por ejemplo, la siguiente condición de combinación devuelve un error de sintaxis:

  ```
  event.eventid*10(+)=category.catid
  ```

  No obstante, la siguiente condición de combinación es válida:

  ```
  event.eventid(+)*10=category.catid
  ```
- No se puede usar el operador (+) en un bloque de consulta que también contiene una sintaxis de combinación de la cláusula FROM.
- Si se combinan dos tablas sobre varias condiciones de combinación, debe usar el operador (+) en todas o ninguna de las siguientes condiciones. Una combinación con estilos de sintaxis mixtos se ejecuta como una combinación interna, sin advertencia.
- El operador (+) no produce una combinación externa si se combina una tabla de la consulta externa con una tabla derivada de una consulta interna.
- Para usar el operador (+) para combinar de manera externa una tabla consigo misma, debe definir los alias de tabla en la cláusula FROM y hacer referencia a ellos en la condición de combinación:

  ```
  select count(*)
  from event a, event b
  where a.eventid(+)=b.catid;

  count
  -------
  8798
  (1 row)
  ```

- No se puede combinar una condición de combinación que contiene el operador (+) con una condición OR o una condición IN. Por ejemplo:

```
select count(*) from sales, listing
where sales.listid(+)=listing.listid or sales.salesid=0;
ERROR:  Outer join operator (+) not allowed in operand of OR or IN.
```

- En la cláusula WHERE que combina de manera externa más de dos tablas, el operador (+) puede aplicarse solo una vez en determinada tabla. En el siguiente ejemplo, no se puede hacer referencia a la tabla SALES con el operador (+) en dos combinaciones consecutivas.

```
select count(*) from sales, listing, event
where sales.listid(+)=listing.listid and sales.dateid(+)=date.dateid;
ERROR:  A table may be outer joined to at most one other table.
```

- Si la condición de combinación externa de la cláusula WHERE compara una columna de la TABLE2 con una constante, aplique el operador (+) a la columna. Si no se incluye el operador, se eliminan las filas combinadas de manera externa de TABLE1, que contienen valores nulos para la columna restringida. Consulte la sección de ejemplos a continuación.

Ejemplos

La siguiente consulta de combinación especifica una combinación externa izquierda de las tablas SALES y LISTING sobre sus columnas LISTID:

```
select count(*)
from sales, listing
where sales.listid = listing.listid(+);

count
--------
172456
(1 row)
```

La siguiente consulta equivalente produce el mismo resultado pero utiliza la sintaxis de combinación de la cláusula FROM:

```
select count(*)
from sales left outer join listing on sales.listid = listing.listid;
```

```
count
--------
172456
(1 row)
```

La tabla SALES no contiene registros de todas las listas de la tabla LISTING, ya que no todas las listas generan ventas. La siguiente consulta combina de manera externa SALES y LISTING, y devuelve filas de LISTING incluso cuando la tabla SALES no informa ventas para un determinado ID de lista. Las columnas PRICE y COMM, que derivan de la tabla SALES, contienen valores nulos en el conjunto de resultados para esas filas no coincidentes.

```
select listing.listid, sum(pricepaid) as price,
sum(commission) as comm
from listing, sales
where sales.listid(+) = listing.listid and listing.listid between 1 and 5
group by 1 order by 1;

listid | price  |  comm
--------+--------+--------
1 | 728.00 | 109.20
2 |        |
3 |        |
4 |  76.00 |  11.40
5 | 525.00 |  78.75
(5 rows)
```

Tenga en cuenta que, cuando se utiliza un operador de combinación de la cláusula WHERE, el orden de las tablas en la cláusula FROM no es importante.

Un ejemplo de una condición de combinación externa más compleja de la cláusula WHERE es el caso en el que la condición consta de una comparación entre dos columnas de tablas y una comparación con una constante:

```
where category.catid=event.catid(+) and eventid(+)=796;
```

Tenga en cuenta que el operador (+) se utiliza en dos lugares: primero, en la comparación de igualdad entre las tablas y, segundo, en la condición de comparación para la columna EVENTID. El resultado de esta sintaxis es la preservación de las filas combinadas de manera externa cuando se evalúa la restricción de EVENTID. Si elimina el operador (+) de la restricción de EVENTID, la consulta trata esta restricción como un filtro, no como parte de la condición de combinación externa.

A su vez, las filas combinadas de manera externa que contienen valores nulos para EVENTID se eliminan del conjunto de resultados.

A continuación se muestra una consulta completa que ilustra este comportamiento:

```
select catname, catgroup, eventid
from category, event
where category.catid=event.catid(+) and eventid(+)=796;

catname | catgroup | eventid
-----------+----------+---------
Classical | Concerts |
Jazz | Concerts |
MLB | Sports   |
MLS | Sports   |
Musicals | Shows    | 796
NBA | Sports   |
NFL | Sports   |
NHL | Sports   |
Opera | Shows    |
Plays | Shows    |
Pop | Concerts |
(11 rows)
```

La consulta equivalente a través de la sintaxis de la cláusula FROM es de la siguiente manera:

```
select catname, catgroup, eventid
from category left join event
on category.catid=event.catid and eventid=796;
```

Si elimina el segundo operador (+) de la versión de la cláusula WHERE de esta consulta, devuelve solo 1 fila (la fila donde `eventid`=796).

```
select catname, catgroup, eventid
from category, event
where category.catid=event.catid(+) and eventid=796;

catname | catgroup | eventid
-----------+----------+---------
Musicals | Shows    | 796
(1 row)
```

## Cláusula GROUP BY

La cláusula GROUP BY identifica las columnas de agrupación para la consulta. Se utiliza para agrupar las filas de una tabla que tienen los mismos valores en todas las columnas de la lista. No importa el orden en el que aparecen las columnas. El resultado es la combinación de cada conjunto de filas que tienen valores comunes en una fila de grupo que representa todas las filas del grupo. Utilice GROUP BY para eliminar la redundancia en la salida y calcular los agregados que se aplican a los grupos. Las columnas de agrupación deben declararse cuando la consulta computa las agregaciones con funciones estándar como SUM, AVG y COUNT. Para obtener más información, consulte Funciones de agregación.

### Sintaxis

```
[ GROUP BY  expression [, ...] | ALL | aggregation_extension  ]
```

donde aggregation_extension es una de las siguientes opciones:

```
GROUPING SETS ( () | aggregation_extension [, ...] ) |
ROLLUP ( expr [, ...] ) |
CUBE ( expr [, ...] )
```

### Parámetros

expression

La lista de columnas o expresiones debe coincidir con la lista de expresiones no agregadas en la lista de selección de la consulta. Por ejemplo, considere la siguiente consulta simple.

```
select listid, eventid, sum(pricepaid) as revenue,
count(qtysold) as numtix
from sales
group by listid, eventid
order by 3, 4, 2, 1
limit 5;

listid | eventid | revenue | numtix
-------+---------+---------+--------
89397  |      47 |   20.00 |      1
106590 |      76 |   20.00 |      1
124683 |     393 |   20.00 |      1
103037 |     403 |   20.00 |      1
```

```
147685 |     429 |   20.00 |      1
(5 rows)
```

En esta consulta, la lista de selección consta de dos expresiones agregadas. La primera usa la función SUM y la segunda usa la función COUNT. Las dos columnas restantes, LISTID y EVENTID, deben declararse como columnas de agrupación.

Las expresiones de la cláusula GROUP BY también pueden hacer referencia a la lista de selección a través de números ordinales. Por ejemplo, el caso anterior podría abreviarse de la siguiente manera.

```
select listid, eventid, sum(pricepaid) as revenue,
count(qtysold) as numtix
from sales
group by 1,2
order by 3, 4, 2, 1
limit 5;

listid | eventid | revenue | numtix
-------+---------+---------+--------
89397  |      47 |   20.00 |      1
106590 |      76 |   20.00 |      1
124683 |     393 |   20.00 |      1
103037 |     403 |   20.00 |      1
147685 |     429 |   20.00 |      1
(5 rows)
```

ALL

ALL indica que se deben agrupar por todas las columnas especificadas en la lista SELECT, excepto las que están agregadas. Por ejemplo, considere la siguiente consulta, que agrupa por `col1` y `col2` sin tener que especificarlas individualmente en la cláusula GROUP BY. La columna `col3` es el argumento de la función SUM y, por lo tanto, no está agrupada.

```
SELECT col1, col2 sum(col3) FROM testtable GROUP BY ALL
```

Si aplica EXCLUDE a una columna de la lista SELECT, la cláusula GROUP BY ALL no agrupa los resultados en función de esa columna específica.

```
SELECT * EXCLUDE col3 FROM testtable GROUP BY ALL
```

aggregation_extension

Puede utilizar las extensiones de agregación GROUPING SETS, ROLLUP y CUBE para realizar el trabajo de varias operaciones GROUP BY en una sola instrucción. Para obtener más información sobre las extensiones de agregación y las funciones relacionadas, consulte Extensiones de agregación.

Ejemplos

Los ejemplos siguientes utilizan la tabla SALES que contiene las columnas: salesid, listid, sellerid, buyerid, eventid, dateid, qtysold, pricepaid, commission y saletime. Para obtener más información sobre la tabla SALES, consulte Base de datos de muestra.

En el siguiente ejemplo, se agrupan consultas por `salesid` y `listid` sin tener que especificarlas individualmente en la cláusula GROUP BY. La columna `qtysold` es el argumento de la función SUM y, por lo tanto, no está agrupada.

```
SELECT salesid, listid, sum(qtysold) FROM sales GROUP BY ALL;

salesid | listid  | sum
--------+---------+------
33095   | 36572   | 2
88268   | 100813  | 4
110917  | 127048  | 1
...
```

El siguiente ejemplo de consulta excluye varias columnas de la lista SELECT, por lo que GROUP BY ALL solo agrupa salesid y listid.

```
SELECT * EXCLUDE sellerid, buyerid, eventid, dateid, qtysold, pricepaid, commission,
 saletime
FROM sales GROUP BY ALL;

salesid | listid
--------+---------
33095   | 36572
88268   | 100813
110917  | 127048
...
```

## Extensiones de agregación

Amazon Redshift admite extensiones de agregación para realizar el trabajo de varias operaciones GROUP BY en una sola instrucción.

Los ejemplos de extensiones de agregación utilizan la tabla `orders`, que contiene los datos de ventas de una empresa de electrónica. Para crear `orders`, puede realizar lo siguiente.

```
CREATE TABLE ORDERS (
    ID INT,
    PRODUCT CHAR(20),
    CATEGORY CHAR(20),
    PRE_OWNED CHAR(1),
    COST DECIMAL
);

INSERT INTO ORDERS VALUES
    (0, 'laptop',       'computers',    'T', 1000),
    (1, 'smartphone',   'cellphones',   'T', 800),
    (2, 'smartphone',   'cellphones',   'T', 810),
    (3, 'laptop',       'computers',    'F', 1050),
    (4, 'mouse',        'computers',    'F', 50);
```

### GROUPING SETS

Calcula uno o más conjuntos de agrupación en una sola instrucción. Un conjunto de agrupación es el conjunto de una sola cláusula GROUP BY, un conjunto de 0 o más columnas mediante el que se puede agrupar el conjunto de resultados de una consulta. GROUP BY GROUPING SETS equivale a ejecutar una consulta UNION ALL en un conjunto de resultados agrupado por columnas diferentes. Por ejemplo, GROUP BY GROUPING SETS((a), (b)) equivale a GROUP BY a UNION ALL GROUP BY b.

En el siguiente ejemplo se devuelve el costo de los productos de la tabla de pedidos agrupados en función tanto de las categorías de los productos como del tipo de productos vendidos.

```


SELECT category, product, sum(cost) as total
FROM orders
GROUP BY GROUPING SETS(category, product);

       category       |       product        | total
----------------------+----------------------+-------
```

```
 computers            |                      |  2100
 cellphones           |                      |  1610
                      | laptop               |  2050
                      | smartphone           |  1610
                      | mouse                |    50

(5 rows)
```

### ROLLUP

Se supone una jerarquía en la que las columnas anteriores se consideran las principales de las columnas posteriores. ROLLUP agrupa los datos por las columnas proporcionadas y devuelve filas de subtotales adicionales que representan los totales de todos los niveles de agrupación de columnas, además de las filas agrupadas. Por ejemplo, puede usar GROUP BY ROLLUP ((a), (b)) para devolver un conjunto de resultados agrupado primero por a y luego por b, suponiendo que b es una subsección de a. ROLLUP también devuelve una fila con todo el conjunto de resultados sin agrupar columnas.

GROUP BY ROLLUP((a), (b)) equivale a GROUP BY GROUPING SETS((a,b), (a), ()).

En el siguiente ejemplo se devuelve el costo de los productos de la tabla de pedidos agrupados primero por categoría y, a continuación, por producto, con el producto como una subdivisión de la categoría.

```
SELECT category, product, sum(cost) as total
FROM orders
GROUP BY ROLLUP(category, product) ORDER BY 1,2;

       category       |       product        | total
----------------------+----------------------+-------
 cellphones           | smartphone           |  1610
 cellphones           |                      |  1610
 computers            | laptop               |  2050
 computers            | mouse                |    50
 computers            |                      |  2100
                      |                      |  3710
(6 rows)
```

### CUBE

Agrupa los datos por las columnas proporcionadas y devuelve filas de subtotales adicionales que representan los totales de todos los niveles de agrupación de columnas, además de las filas

agrupadas. CUBE devuelve las mismas filas que ROLLUP, a la vez que agrega filas de subtotales adicionales por cada combinación de columnas de agrupación no incluidas en ROLLUP. Por ejemplo, puede usar GROUP BY CUBE ((a), (b)) para devolver un conjunto de resultados agrupado primero por a y luego por b, suponiendo que b es una subsección de a. CUBE también devuelve una fila con todo el conjunto de resultados sin agrupar columnas.

GROUP BY CUBE((a), (b)) equivale a GROUP BY GROUPING SETS((a, b), (a), (b), ()).

En el siguiente ejemplo se devuelve el costo de los productos de la tabla de pedidos agrupados primero por categoría y, a continuación, por producto, con el producto como una subdivisión de la categoría. A diferencia del ejemplo anterior de ROLLUP, la instrucción devuelve resultados para cada combinación de columnas de agrupación.

```
SELECT category, product, sum(cost) as total
FROM orders
GROUP BY CUBE(category, product) ORDER BY 1,2;

       category       |       product        | total
----------------------+----------------------+-------
 cellphones           | smartphone           |  1610
 cellphones           |                      |  1610
 computers            | laptop               |  2050
 computers            | mouse                |    50
 computers            |                      |  2100
                      | laptop               |  2050
                      | mouse                |    50
                      | smartphone           |  1610
                      |                      |  3710
(9 rows)
```

Funciones de GROUPING/GROUPING_ID

ROLLUP y CUBE agregan valores NULL al conjunto de resultados para indicar las filas del subtotal. Por ejemplo, GROUP BY ROLLUP((a), (b)) devuelve una o varias filas que tienen un valor NULL en la columna de agrupación b para indicar que son subtotales de campos de la columna de agrupación a. Estos valores NULL solo sirven para satisfacer el formato de las tuplas devueltas.

Al ejecutar las operaciones GROUP BY con ROLLUP y CUBE en relaciones que almacenan valores NULL por sí mismas, se pueden producir conjuntos de resultados con filas que parecen tener columnas de agrupación idénticas. Volviendo al ejemplo anterior, si la columna de agrupación b

contiene un valor NULL almacenado, GROUP BY ROLLUP((a), (b)) devuelve una fila con un valor NULL en la columna de agrupación b que no es un subtotal.

Para distinguir entre los valores NULL creados por ROLLUP y CUBE, y los valores NULL almacenados en las propias tablas, puede utilizar la función GROUPING o su alias GROUPING_ID. GROUPING toma un único conjunto de agrupación como argumento y, para cada fila del conjunto de resultados, devuelve un valor de bit 0 o 1 correspondiente a la columna de agrupación en esa posición y, a continuación, convierte el valor en un entero. Si el valor de esa posición es un valor NULL creado por una extensión de agregación, GROUPING devuelve 1. Devuelve 0 para todos los demás valores, incluidos los valores NULL almacenados.

Por ejemplo, GROUPING(category, product) puede devolver los siguientes valores para una fila determinada, en función de los valores de la columna de agrupación de esa fila. Para los fines de este ejemplo, todos los valores NULL de la tabla son valores NULL creados por una extensión de agregación.

| Columna category | Columna product | Valor de bit de la función GROUPING | Valor decimal |
|---|---|---|---|
| no NULL | no NULL | 00 | 0 |
| no NULL | NULL | 01 | 1 |
| NULL | no NULL | 10 | 2 |
| NULL | NULL | 11 | 3 |

Las funciones de GROUPING aparecen en la parte de la lista SELECT de la consulta en el siguiente formato.

```
SELECT ... [GROUPING( expr )...] ...
  GROUP BY ... {CUBE | ROLLUP| GROUPING SETS} ( expr ) ...
```

El siguiente ejemplo es el mismo que el anterior de CUBE, pero con la adición de funciones de GROUPING para sus conjuntos de agrupación.

```
SELECT category, product,
       GROUPING(category) as grouping0,
```

```
       GROUPING(product) as grouping1,
       GROUPING(category, product) as grouping2,
       sum(cost) as total
FROM orders
GROUP BY CUBE(category, product) ORDER BY 3,1,2;

       category       |       product        | grouping0 | grouping1 | grouping2 |
 total
----------------------+----------------------+-----------+-----------+-----------
+-------
 cellphones           | smartphone           |         0 |         0 |         0 |
 1610
 cellphones           |                      |         0 |         1 |         1 |
 1610
 computers            | laptop               |         0 |         0 |         0 |
 2050
 computers            | mouse                |         0 |         0 |         0 |
 50
 computers            |                      |         0 |         1 |         1 |
 2100
                      | laptop               |         1 |         0 |         2 |
 2050
                      | mouse                |         1 |         0 |         2 |
 50
                      | smartphone           |         1 |         0 |         2 |
 1610
                      |                      |         1 |         1 |         3 |
 3710
(9 rows)
```

ROLLUP y CUBE parciales

Puede ejecutar las operaciones ROLLUP y CUBE con solo una parte de los subtotales.

La sintaxis de las operaciones ROLLUP y CUBE parciales es la siguiente.

```
GROUP BY expr1, { ROLLUP | CUBE }(expr2, [, ...])
```

En este caso, la cláusula GROUP BY solo crea filas de subtotales en el nivel de expr2 y posteriores.

En los siguientes ejemplos se muestran operaciones parciales de ROLLUP y CUBE en la tabla orders. Primero se agrupan si un producto es de segunda mano y, a continuación, se ejecutan ROLLUP y CUBE en las columnas category y product.

```
SELECT pre_owned, category, product,
       GROUPING(category, product, pre_owned) as group_id,
       sum(cost) as total
FROM orders
GROUP BY pre_owned, ROLLUP(category, product) ORDER BY 4,1,2,3;

 pre_owned |       category       |       product        | group_id | total
-----------+----------------------+----------------------+----------+-------
 F         | computers            | laptop               |        0 |  1050
 F         | computers            | mouse                |        0 |    50
 T         | cellphones           | smartphone           |        0 |  1610
 T         | computers            | laptop               |        0 |  1000
 F         | computers            |                      |        2 |  1100
 T         | cellphones           |                      |        2 |  1610
 T         | computers            |                      |        2 |  1000
 F         |                      |                      |        6 |  1100
 T         |                      |                      |        6 |  2610
(9 rows)

SELECT pre_owned, category, product,
       GROUPING(category, product, pre_owned) as group_id,
       sum(cost) as total
FROM orders
GROUP BY pre_owned, CUBE(category, product) ORDER BY 4,1,2,3;

 pre_owned |       category       |       product        | group_id | total
-----------+----------------------+----------------------+----------+-------
 F         | computers            | laptop               |        0 |  1050
 F         | computers            | mouse                |        0 |    50
 T         | cellphones           | smartphone           |        0 |  1610
 T         | computers            | laptop               |        0 |  1000
 F         | computers            |                      |        2 |  1100
 T         | cellphones           |                      |        2 |  1610
 T         | computers            |                      |        2 |  1000
 F         |                      | laptop               |        4 |  1050
 F         |                      | mouse                |        4 |    50
 T         |                      | laptop               |        4 |  1000
 T         |                      | smartphone           |        4 |  1610
 F         |                      |                      |        6 |  1100
 T         |                      |                      |        6 |  2610
(13 rows)
```

Como la columna de segunda mano (pre_owned) no se incluye en las operaciones ROLLUP y CUBE, no existe una fila de total general que incluya todas las demás filas.

Agrupación concatenada

Puede concatenar varias cláusulas GROUPING SETS/ROLLUP/CUBE para calcular diferentes niveles de subtotales. Las agrupaciones concatenadas devuelven el producto cartesiano de los conjuntos de agrupación proporcionados.

La sintaxis para concatenar cláusulas GROUPING SETS/ROLLUP/CUBE es la siguiente.

```
GROUP BY {ROLLUP|CUBE|GROUPING SETS}(expr1[, ...]),
         {ROLLUP|CUBE|GROUPING SETS}(expr1[, ...])[, ...]
```

Considere el siguiente ejemplo para ver cómo una agrupación concatenada pequeña puede producir un gran conjunto de resultados finales.

```
SELECT pre_owned, category, product,
       GROUPING(category, product, pre_owned) as group_id,
       sum(cost) as total
FROM orders
GROUP BY CUBE(category, product), GROUPING SETS(pre_owned, ())
ORDER BY 4,1,2,3;

 pre_owned |       category       |       product        | group_id | total
-----------+----------------------+----------------------+----------+-------
 F         | computers            | laptop               |        0 |  1050
 F         | computers            | mouse                |        0 |    50
 T         | cellphones           | smartphone           |        0 |  1610
 T         | computers            | laptop               |        0 |  1000
           | cellphones           | smartphone           |        1 |  1610
           | computers            | laptop               |        1 |  2050
           | computers            | mouse                |        1 |    50
 F         | computers            |                      |        2 |  1100
 T         | cellphones           |                      |        2 |  1610
 T         | computers            |                      |        2 |  1000
           | cellphones           |                      |        3 |  1610
           | computers            |                      |        3 |  2100
 F         |                      | laptop               |        4 |  1050
 F         |                      | mouse                |        4 |    50
 T         |                      | laptop               |        4 |  1000
 T         |                      | smartphone           |        4 |  1610
```

```
           |                      | laptop               |        5 |  2050
           |                      | mouse                |        5 |    50
           |                      | smartphone           |        5 |  1610
 F         |                      |                      |        6 |  1100
 T         |                      |                      |        6 |  2610
           |                      |                      |        7 |  3710
(22 rows)
```

Agrupación anidada

Puede utilizar las operaciones GROUPING SETS/ROLLUP/CUBE como GROUPING SETS expr para formar una agrupación anidada. La subagrupación dentro de GROUPING SETS anidado está aplanada.

La sintaxis de la agrupación anidada es la siguiente.

```
GROUP BY GROUPING SETS({ROLLUP|CUBE|GROUPING SETS}(expr[, ...])[, ...])
```

Considere el siguiente ejemplo.

```
SELECT category, product, pre_owned,
       GROUPING(category, product, pre_owned) as group_id,
       sum(cost) as total
FROM orders
GROUP BY GROUPING SETS(ROLLUP(category), CUBE(product, pre_owned))
ORDER BY 4,1,2,3;

       category       |       product        | pre_owned | group_id | total
----------------------+----------------------+-----------+----------+-------
 cellphones           |                      |           |        3 |  1610
 computers            |                      |           |        3 |  2100
                      | laptop               | F         |        4 |  1050
                      | laptop               | T         |        4 |  1000
                      | mouse                | F         |        4 |    50
                      | smartphone           | T         |        4 |  1610
                      | laptop               |           |        5 |  2050
                      | mouse                |           |        5 |    50
                      | smartphone           |           |        5 |  1610
                      |                      | F         |        6 |  1100
                      |                      | T         |        6 |  2610
                      |                      |           |        7 |  3710
                      |                      |           |        7 |  3710
```

```
(13 rows)
```

Tenga en cuenta que, como ROLLUP(category) y CUBE(product, pre_owned) contienen el conjunto de agrupación (), la fila que representa el total general está duplicada.

Notas de uso

- La cláusula GROUP BY admite hasta 64 conjuntos de agrupación. En el caso de ROLLUP y CUBE, o de alguna combinación de GROUPING SETS, ROLLUP y CUBE, esta limitación se aplica al número implícito de conjuntos de agrupación. Por ejemplo, GROUP BY CUBE((a), (b)) cuenta como cuatro conjuntos de agrupación, no como dos.
- No puede utilizar constantes como columnas de agrupación cuando utilice extensiones de agregación.
- No puede hacer un conjunto de agrupación que contenga columnas duplicadas.

## Cláusula HAVING

La cláusula HAVING aplica una condición al conjunto de resultados agrupado intermedio que una consulta devuelve.

Sintaxis

```
[ HAVING condition ]
```

Por ejemplo, puede limitar los resultados de una función SUM:

```
having sum(pricepaid) >10000
```

La condición HAVING se aplica después de que se aplican todas las condiciones de la cláusula WHERE y se completan todas las operaciones de GROUP BY.

La condición toma la misma forma que cualquier condición de la cláusula WHERE.

Notas de uso

- Cualquier columna a la que se haga referencia en una condición de la cláusula HAVING debe ser una columna de agrupación o una columna que haga referencia al resultado de una función agregada.
- En una cláusula HAVING, no se puede especificar:

- Un número ordinal que hace referencia a un elemento de la lista de selección. Solo las cláusulas GROUP BY y ORDER BY aceptan números ordinales.

Ejemplos

La siguiente consulta calcula las ventas de tickets totales para todos los eventos por nombre y, luego, elimina eventos donde las ventas totales sean inferiores a $800 000. La condición HAVING se aplica a los resultados de la función agregada en la lista de selección: `sum(pricepaid)`.

```
select eventname, sum(pricepaid)
from sales join event on sales.eventid = event.eventid
group by 1
having sum(pricepaid) > 800000
order by 2 desc, 1;

eventname        |    sum
-----------------+-----------
Mamma Mia!       | 1135454.00
Spring Awakening |  972855.00
The Country Girl |  910563.00
Macbeth          |  862580.00
Jersey Boys      |  811877.00
Legally Blonde   |  804583.00
```

La siguiente consulta calcula un conjunto de resultados similar. No obstante, en este caso, la condición HAVING se aplica a una agregación que no se especifica en la lista de selección: `sum(qtysold)`. Los eventos que no vendieron más de 2 000 tickets se eliminan del resultado final.

```
select eventname, sum(pricepaid)
from sales join event on sales.eventid = event.eventid
group by 1
having sum(qtysold) >2000
order by 2 desc, 1;

eventname        |    sum
-----------------+-----------
Mamma Mia!       | 1135454.00
Spring Awakening |  972855.00
The Country Girl |  910563.00
Macbeth          |  862580.00
Jersey Boys      |  811877.00
```

```
Legally Blonde   |  804583.00
Chicago          |  790993.00
Spamalot         |  714307.00
```

La siguiente consulta calcula las ventas de tickets totales para todos los eventos por nombre y, luego, elimina eventos donde las ventas totales sean inferiores a $800 000. La condición HAVING se aplica a los resultados de la función agregada en la lista de selección utilizando el alias pp para `sum(pricepaid)`.

```
select eventname, sum(pricepaid) as pp
from sales join event on sales.eventid = event.eventid
group by 1
having pp > 800000
order by 2 desc, 1;

eventname        |    pp
-----------------+-----------
Mamma Mia!       | 1135454.00
Spring Awakening |  972855.00
The Country Girl |  910563.00
Macbeth          |  862580.00
Jersey Boys      |  811877.00
Legally Blonde   |  804583.00
```

## Cláusula QUALIFY

La cláusula QUALIFY filtra los resultados de una función de ventana calculada previamente en función de las condiciones de búsqueda especificadas por el usuario. Puede utilizar la cláusula para aplicar condiciones de filtrado al resultado de una función de ventana sin utilizar una subconsulta.

Es similar a la cláusula HAVING, que aplica una condición para filtrar aún más las filas de una cláusula WHERE. La diferencia entre QUALIFY y HAVING es que los resultados filtrados de la cláusula QUALIFY se pueden basar en el resultado de ejecutar funciones de ventana en los datos. Puede utilizar las cláusulas QUALITY y HAVING en una consulta.

### Sintaxis

```
QUALIFY condition
```

> **Note**
>
> Si utiliza la cláusula QUALIFY directamente después de la cláusula FROM, el nombre de la relación FROM debe tener un alias especificado antes de la cláusula QUALIFY.

### Ejemplos

En los ejemplos de esta sección se utilizan los datos de ejemplo que se muestran a continuación.

```
create table store_sales (ss_sold_date date, ss_sold_time time,
               ss_item text, ss_sales_price float);
insert into store_sales values ('2022-01-01', '09:00:00', 'Product 1', 100.0),
                               ('2022-01-01', '11:00:00', 'Product 2', 500.0),
                               ('2022-01-01', '15:00:00', 'Product 3', 20.0),
                               ('2022-01-01', '17:00:00', 'Product 4', 1000.0),
                               ('2022-01-01', '18:00:00', 'Product 5', 30.0),
                               ('2022-01-02', '10:00:00', 'Product 6', 5000.0),
                               ('2022-01-02', '16:00:00', 'Product 7', 5.0);
```

El siguiente ejemplo muestra cómo encontrar los dos artículos más caros que se venden después de las 12:00 cada día.

```
SELECT *
FROM store_sales ss
WHERE ss_sold_time > time '12:00:00'
QUALIFY row_number()
OVER (PARTITION BY ss_sold_date ORDER BY ss_sales_price DESC) <= 2


 ss_sold_date | ss_sold_time |  ss_item  | ss_sales_price
--------------+--------------+-----------+----------------
 2022-01-01   | 17:00:00     | Product 4 |           1000
 2022-01-01   | 18:00:00     | Product 5 |             30
 2022-01-02   | 16:00:00     | Product 7 |              5
```

A continuación, podrá encontrar el último artículo vendido cada día.

```
SELECT *
FROM store_sales ss
QUALIFY last_value(ss_item)
OVER (PARTITION BY ss_sold_date ORDER BY ss_sold_time ASC
```

```
      ROWS BETWEEN UNBOUNDED PRECEDING AND UNBOUNDED FOLLOWING) = ss_item;

ss_sold_date | ss_sold_time |  ss_item  | ss_sales_price
-------------+--------------+-----------+---------------
 2022-01-01   | 18:00:00     | Product 5 |             30
 2022-01-02   | 16:00:00     | Product 7 |              5
```

El siguiente ejemplo devuelve los mismos registros que la consulta anterior, el último artículo vendido cada día, pero no utiliza la cláusula QUALIFY.

```
SELECT * FROM (
  SELECT *,
  last_value(ss_item)
  OVER (PARTITION BY ss_sold_date ORDER BY ss_sold_time ASC
        ROWS BETWEEN UNBOUNDED PRECEDING AND UNBOUNDED FOLLOWING) ss_last_item
  FROM store_sales ss
)
WHERE ss_last_item = ss_item;

 ss_sold_date | ss_sold_time |  ss_item  | ss_sales_price | ss_last_item
-------------+--------------+-----------+----------------+--------------
 2022-01-02   | 16:00:00     | Product 7 |              5 | Product 7
 2022-01-01   | 18:00:00     | Product 5 |             30 | Product 5
```

## UNION, INTERSECT y EXCEPT

Temas

- Sintaxis
- Parámetros
- Orden de evaluación para los operadores de conjunto
- Notas de uso
- Ejemplo de consultas UNION
- Ejemplo de consultas UNION ALL
- Ejemplo de consultas INTERSECT
- Ejemplo de consulta EXCEPT

Los operadores de conjunto UNION, INTERSECT y EXCEPT se utilizan para comparar y fusionar los resultados de dos expresiones de consulta diferentes. Por ejemplo, si desea saber qué usuarios

de un sitio web son compradores y vendedores pero los nombres de usuario están almacenados en diferentes columnas o tablas, puede buscar la intersección de estos dos tipos de usuarios. Si desea saber qué usuarios de un sitio web son compradores pero no vendedores, puede usar el operador EXCEPT para buscar la diferencia entre las dos listas de usuarios. Si desea crear una lista de todos los usuarios, independientemente de la función, puede usar el operador UNION.

Sintaxis

```
query
{ UNION [ ALL ] | INTERSECT | EXCEPT | MINUS }
query
```

Parámetros

consulta

Una expresión de consulta que corresponde, en la forma de su lista de selección, a una segunda expresión de consulta que sigue el operador UNION, INTERSECT o EXCEPT. Las dos expresiones deben contener la misma cantidad de columnas de salida con tipos de datos compatibles; de lo contrario, no se podrán comparar ni fusionar los dos conjuntos de resultados. Las operaciones de conjunto no permiten conversiones implícitas entre diferentes categorías de tipos de datos. Para obtener más información, consulte Conversión y compatibilidad de tipos.

Puede crear consultas que contengan una cantidad ilimitada de expresiones de consulta y vincularlas con operadores UNION, INTERSECT y EXCEPT en cualquier combinación. Por ejemplo, la siguiente estructura de consulta es válida, suponiendo que las tablas T1, T2 y T3 contienen conjuntos de columnas compatibles:

```
select * from t1
union
select * from t2
except
select * from t3
order by c1;
```

UNION

Operación de conjunto que devuelve filas de dos expresiones de consulta, independientemente de si las filas provienen de una o ambas expresiones.

INTERSECT

Operación de conjunto que devuelve filas que provienen de dos expresiones de consulta. Las filas que no se devuelven en las dos expresiones se descartan.

EXCEPT | MINUS

Operación de conjunto que devuelve filas que provienen de una de las dos expresiones de consulta. Para calificar para el resultado, las filas deben existir en la primera tabla de resultados, pero no en la segunda. MINUS y EXCEPT son sinónimos exactos.

ALL

La palabra clave ALL conserva cualquier fila duplicada que UNION produce. El comportamiento predeterminado cuando no se usa la palabra clave ALL es descartar todos estos duplicados. No se admiten las expresiones INTERSECT ALL, EXCEPT ALL y MINUS ALL.

### Orden de evaluación para los operadores de conjunto

Los operadores de conjunto UNION y EXCEPT se asocian por la izquierda. Si no se especifican paréntesis para establecer el orden de prioridad, los operadores se evalúan de izquierda a derecha. Por ejemplo, en la siguiente consulta, UNION de T1 y T2 se evalúa primero, luego se realiza la operación EXCEPT en el resultado de UNION:

```
select * from t1
union
select * from t2
except
select * from t3
order by c1;
```

El operador INTERSECT prevalece sobre los operadores UNION y EXCEPT cuando se utiliza una combinación de operadores en la misma consulta. Por ejemplo, la siguiente consulta evalúa la intersección de T2 y T3, y luego unirá el resultado con T1:

```
select * from t1
union
select * from t2
intersect
select * from t3
order by c1;
```

Al agregar paréntesis, puede aplicar un orden diferente de evaluación. En el siguiente caso, el resultado de la unión de T1 y T2 está intersectado con T3, y la consulta probablemente produzca un resultado diferente.

```
(select * from t1
union
select * from t2)
intersect
(select * from t3)
order by c1;
```

Notas de uso

- Los nombres de la columnas que se devuelven en el resultado de una consulta de operación de conjunto son los nombres (o alias) de la columnas de las tablas de la primera expresión de consulta. Debido a que estos nombres de columnas pueden ser confusos, porque los valores de la columna provienen de tablas de cualquier lado del operador de conjunto, se recomienda proporcionar alias significativos para el conjunto de resultados.
- Una expresión de consulta que precede a un operador de conjunto no debería contener una cláusula ORDER BY. Una cláusula ORDER BY produce resultados significativos ordenados solo cuando se utiliza al final de una consulta que contiene operadores de conjunto. En este caso, la cláusula ORDER BY se aplica a los resultados finales de todas las operaciones de conjunto. La consulta extrema también puede contener cláusula LIMIT y OFFSET estándar.
- Cuando las consultas del operador de conjunto devuelven resultados decimales, las columnas de resultado correspondientes se promueven a devolver la misma precisión y escala. Por ejemplo, en la siguiente consulta, donde T1.REVENUE es una columna DECIMAL(10,2) y T2.REVENUE es una columna DECIMAL(8,4), el resultado decimal se promueve a DECIMAL(12,4):

  ```
  select t1.revenue union select t2.revenue;
  ```

  La escala es 4 ya que es la escala máxima de las dos columnas. La precisión es 12 ya que T1.REVENUE requiere 8 dígitos a la izquierda del punto decimal (12 - 4 = 8). Este tipo de promoción garantiza que todos los valores de ambos lados de UNION encajen en el resultado. Para valores de 64 bits, la precisión de resultados máxima es 19 y la escala de resultados máxima es 18. Para valores de 128 bits, la precisión de resultados máxima es 38 y la escala de resultados máxima es 37.

Si el tipo de datos resultante supera los límites de precisión y escala de Amazon Redshift, la consulta devuelve un error.

- En el caso de las operaciones de conjunto, las dos filas se tratan como idénticas si, para cada par de columnas correspondiente, los dos valores de datos son iguales o NULL. Por ejemplo, si las tablas T1 y T2 contienen una columna y una fila, y esa fila es NULL en ambas tablas, una operación INTERSECT sobre esas tablas devuelve esa fila.

### Ejemplo de consultas UNION

En la siguiente consulta UNION, las filas de la tabla SALES se fusionan con las filas de la tabla LISTING. Se seleccionan tres columnas compatibles de cada tabla. En este caso, las columnas correspondientes tienen los mismos nombres y tipos de datos.

El conjunto de resultados finales está ordenado por la primera columna de la tabla LISTING y limitado a las 5 filas con el mayor valor LISTID.

```
select listid, sellerid, eventid from listing
union select listid, sellerid, eventid from sales
order by listid, sellerid, eventid desc limit 5;

listid | sellerid | eventid
--------+----------+---------
1 |    36861 |    7872
2 |    16002 |    4806
3 |    21461 |    4256
4 |     8117 |    4337
5 |     1616 |    8647
(5 rows)
```

En el siguiente ejemplo, se muestra cómo puede agregar un valor literal para el resultado de una consulta UNION para ver cuál expresión de consulta produjo cada fila en el conjunto de resultados. La consulta identifica filas de la primera expresión de consulta como "B" (por compradores, "buyers" en inglés) y filas de la segunda expresión de consulta como "S" (por vendedores, "sellers" en inglés).

La consulta identifica compradores y vendedores para transacciones de ticket que cuestan $10 000 o más. La única diferencia entre las dos expresiones de consulta de cualquier lado del operador UNION es la columna de combinación para la tabla SALES.

```
select listid, lastname, firstname, username,
```

```
pricepaid as price, 'S' as buyorsell
from sales, users
where sales.sellerid=users.userid
and pricepaid >=10000
union
select listid, lastname, firstname, username, pricepaid,
'B' as buyorsell
from sales, users
where sales.buyerid=users.userid
and pricepaid >=10000
order by 1, 2, 3, 4, 5;

listid | lastname | firstname | username |   price   | buyorsell
--------+----------+-----------+----------+-----------+-----------
209658 | Lamb     | Colette   | VOR15LYI |  10000.00 | B
209658 | West     | Kato      | ELU81XAA |  10000.00 | S
212395 | Greer    | Harlan    | GXO71KOC |  12624.00 | S
212395 | Perry    | Cora      | YWR73YNZ |  12624.00 | B
215156 | Banks    | Patrick   | ZNQ69CLT |  10000.00 | S
215156 | Hayden   | Malachi   | BBG56AKU |  10000.00 | B
(6 rows)
```

En el siguiente ejemplo, se utiliza un operador UNION ALL porque las filas duplicadas, si se encuentran, deben conservarse en el resultado. Para una serie específica de ID de eventos, la consulta devuelve 0 o más filas para cada venta asociada a cada evento, y 0 o 1 fila para cada lista de ese evento. Los ID de eventos son únicos en cada fila de las tablas LISTING y EVENT, pero es posible que haya varias ventas para la misma combinación de ID de lista y evento en la tabla SALES.

La tercera columna en el conjunto de resultados identifica la fuente de la fila. Si viene de la tabla SALES, se marca "Sí" en la columna SALESROW. (SALESROW es un alias para SALES.LISTID). Si la fila proviene de la tabla LISTING, se marca "No" en la columna SALESROW.

En este caso, el conjunto de resultados consta de tres filas de ventas para la lista 500, evento 7787. En otras palabras, se llevaron a cabo tres transacciones diferentes para esta combinación de lista y evento. Las otras dos listas, 501 y 502, no produjeron ventas, por lo que la única fila que la consulta produce para estos ID de lista provienen de la tabla LISTING (SALESROW = 'No').

```
select eventid, listid, 'Yes' as salesrow
from sales
where listid in(500,501,502)
union all
select eventid, listid, 'No'
```

```
from listing
where listid in(500,501,502)
order by listid asc;

eventid | listid | salesrow
---------+--------+----------
7787 |    500 | No
7787 |    500 | Yes
7787 |    500 | Yes
7787 |    500 | Yes
6473 |    501 | No
5108 |    502 | No
(6 rows)
```

Si ejecuta la misma consulta sin la palabra clave ALL, el resultado conserva solo una de las transacciones de ventas.

```
select eventid, listid, 'Yes' as salesrow
from sales
where listid in(500,501,502)
union
select eventid, listid, 'No'
from listing
where listid in(500,501,502)
order by listid asc;

eventid | listid | salesrow
---------+--------+----------
7787 |    500 | No
7787 |    500 | Yes
6473 |    501 | No
5108 |    502 | No
(4 rows)
```

Ejemplo de consultas UNION ALL

En el siguiente ejemplo, se utiliza un operador UNION ALL porque las filas duplicadas, si se encuentran, deben conservarse en el resultado. Para una serie específica de ID de eventos, la consulta devuelve 0 o más filas para cada venta asociada a cada evento, y 0 o 1 fila para cada lista de ese evento. Los ID de eventos son únicos en cada fila de las tablas LISTING y EVENT, pero es posible que haya varias ventas para la misma combinación de ID de lista y evento en la tabla SALES.

La tercera columna en el conjunto de resultados identifica la fuente de la fila. Si viene de la tabla SALES, se marca "Sí" en la columna SALESROW. (SALESROW es un alias para SALES.LISTID). Si la fila proviene de la tabla LISTING, se marca "No" en la columna SALESROW.

En este caso, el conjunto de resultados consta de tres filas de ventas para la lista 500, evento 7787. En otras palabras, se llevaron a cabo tres transacciones diferentes para esta combinación de lista y evento. Las otras dos listas, 501 y 502, no produjeron ventas, por lo que la única fila que la consulta produce para estos ID de lista provienen de la tabla LISTING (SALESROW = 'No').

```
select eventid, listid, 'Yes' as salesrow
from sales
where listid in(500,501,502)
union all
select eventid, listid, 'No'
from listing
where listid in(500,501,502)
order by listid asc;

eventid | listid | salesrow
---------+--------+----------
7787 |    500 | No
7787 |    500 | Yes
7787 |    500 | Yes
7787 |    500 | Yes
6473 |    501 | No
5108 |    502 | No
(6 rows)
```

Si ejecuta la misma consulta sin la palabra clave ALL, el resultado conserva solo una de las transacciones de ventas.

```
select eventid, listid, 'Yes' as salesrow
from sales
where listid in(500,501,502)
union
select eventid, listid, 'No'
from listing
where listid in(500,501,502)
order by listid asc;

eventid | listid | salesrow
---------+--------+----------
```

```
7787 |    500 | No
7787 |    500 | Yes
6473 |    501 | No
5108 |    502 | No
(4 rows)
```

Ejemplo de consultas INTERSECT

Compare el siguiente ejemplo con el primer ejemplo de UNION. La única diferencia entre los dos ejemplos es el operador de conjunto que se utiliza, pero los resultados son muy diferentes. Solo una de las filas es igual:

```
235494 |    23875 |    8771
```

Esta es la única fila en el resultado limitado de 5 filas que se encontró en ambas tablas.

```
select listid, sellerid, eventid from listing
intersect
select listid, sellerid, eventid from sales
order by listid desc, sellerid, eventid
limit 5;

listid | sellerid | eventid
--------+----------+---------
235494 |    23875 |    8771
235482 |     1067 |    2667
235479 |     1589 |    7303
235476 |    15550 |     793
235475 |    22306 |    7848
(5 rows)
```

La siguiente consulta busca eventos (para los que se vendieron tickets) que ocurrieron en lugares en la Ciudad de Nueva York y Los Ángeles en marzo. La diferencia entre las dos expresiones de consulta es la restricción en la columna VENUECITY.

```
select distinct eventname from event, sales, venue
where event.eventid=sales.eventid and event.venueid=venue.venueid
and date_part(month,starttime)=3 and venuecity='Los Angeles'
intersect
select distinct eventname from event, sales, venue
where event.eventid=sales.eventid and event.venueid=venue.venueid
and date_part(month,starttime)=3 and venuecity='New York City'
```

```
order by eventname asc;

eventname
---------------------------
A Streetcar Named Desire
Dirty Dancing
Electra
Running with Annalise
Hairspray
Mary Poppins
November
Oliver!
Return To Forever
Rhinoceros
South Pacific
The 39 Steps
The Bacchae
The Caucasian Chalk Circle
The Country Girl
Wicked
Woyzeck
(16 rows)
```

Ejemplo de consulta EXCEPT

La tabla CATEGORY de la base de datos TICKIT contiene las siguientes 11 filas:

```
 catid | catgroup |  catname  |                 catdesc
-------+----------+-----------+--------------------------------------------
   1   | Sports   | MLB       | Major League Baseball
   2   | Sports   | NHL       | National Hockey League
   3   | Sports   | NFL       | National Football League
   4   | Sports   | NBA       | National Basketball Association
   5   | Sports   | MLS       | Major League Soccer
   6   | Shows    | Musicals  | Musical theatre
   7   | Shows    | Plays     | All non-musical theatre
   8   | Shows    | Opera     | All opera and light opera
   9   | Concerts | Pop       | All rock and pop music concerts
  10   | Concerts | Jazz      | All jazz singers and bands
  11   | Concerts | Classical | All symphony, concerto, and choir concerts
(11 rows)
```

Supongamos que una tabla CATEGORY_STAGE (una tabla provisional) contiene una fila adicional:

```
 catid | catgroup |  catname  |                 catdesc
-------+----------+-----------+--------------------------------------------
1 | Sports   | MLB       | Major League Baseball
2 | Sports   | NHL       | National Hockey League
3 | Sports   | NFL       | National Football League
4 | Sports   | NBA       | National Basketball Association
5 | Sports   | MLS       | Major League Soccer
6 | Shows    | Musicals  | Musical theatre
7 | Shows    | Plays     | All non-musical theatre
8 | Shows    | Opera     | All opera and light opera
9 | Concerts | Pop       | All rock and pop music concerts
10 | Concerts | Jazz      | All jazz singers and bands
11 | Concerts | Classical | All symphony, concerto, and choir concerts
12 | Concerts | Comedy    | All stand up comedy performances
(12 rows)
```

Devuelve la diferencia entre las dos tablas. En otras palabras, devuelve filas que están en la tabla CATEGORY_STAGE pero no en la tabla CATEGORY:

```
select * from category_stage
except
select * from category;

catid | catgroup | catname |             catdesc
-------+----------+---------+----------------------------------
12 | Concerts | Comedy  | All stand up comedy performances
(1 row)
```

La siguiente consulta equivalente usa el sinónimo MINUS.

```
select * from category_stage
minus
select * from category;

catid | catgroup | catname |             catdesc
-------+----------+---------+----------------------------------
12 | Concerts | Comedy  | All stand up comedy performances
(1 row)
```

Si revierte el orden de las expresiones SELECT, la consulta no devuelve filas.

# Cláusula ORDER BY

Temas


- Sintaxis
- Parámetros
- Notas de uso
- Ejemplos con ORDER BY


La cláusula ORDER BY ordena el conjunto de resultados de una consulta.

## Sintaxis

```
[ ORDER BY expression [ ASC | DESC ] ]
[ NULLS FIRST | NULLS LAST ]
[ LIMIT { count | ALL } ]
[ OFFSET start ]
```

## Parámetros

expression

Expresión que define el orden del conjunto de resultados de la consulta, por lo general al especificar una o más columnas en la lista de selección. Los resultados se devuelven en función de la ordenación UTF-8 binaria. También puede especificar lo siguiente:

- Columnas que no estén en la lista de selección
- Expresiones formadas por una o más columnas que existen en las tablas a las que la consulta hace referencia
- Números ordinales que representan la posición de las entradas de la lista de selección (o la posición de columnas en la tabla si no existe una lista de selección)
- Alias que definen las entradas de la lista de selección

Cuando la cláusula ORDER BY contiene varias expresiones, el conjunto de resultados se ordena según la primera expresión, luego se aplica la segunda expresión a las filas que tienen valores coincidentes de la primera expresión, etc.

ASC | DESC

Opción que define el orden de ordenación para la expresión, de la siguiente manera:

- ASC: ascendente (por ejemplo, de menor a mayor para valores numéricos y de la A a la Z para cadenas con caracteres). Si no se especifica ninguna opción, los datos se ordenan, de manera predeterminada, en orden ascendente.
- DESC: descendente (de mayor a menor para valores numéricos y de la Z a la A para cadenas).

NULLS FIRST | NULLS LAST

Opción que especifica si los valores NULL se deben ordenar en primer lugar, antes de los valores no nulos, o al final, después de los valores no nulos. De manera predeterminada, los valores NULL se ordenan y clasificación al final en orden ASC, y se ordenan y se clasifican primero en orden DESC.

LIMIT number (número) | ALL

Opción que controla la cantidad de filas ordenadas que una consulta devuelve. El número LIMIT deber ser un entero positivo; el valor máximo es 2147483647.

LIMIT 0 no devuelve filas. Puede usar la sintaxis para realizar pruebas: para verificar que una consulta se ejecuta (sin mostrar filas) o para devolver una lista de columnas de una tabla. Una cláusula ORDER BY es redundante si está utilizando LIMIT 0 para devolver una lista de columnas. El predeterminado es LIMIT ALL.

OFFSET start (inicio)

Opción que especifica que se omita el número de filas que hay delante de start (inicio) antes de comenzar a devolver filas. El número OFFSET deber ser un entero positivo; el valor máximo es 2147483647. Cuando se utiliza con la opción LIMIT, las filas OFFSET se omiten antes de comenzar a contar las filas LIMIT que se devuelven. Si no se utiliza la opción LIMIT, la cantidad de filas del conjunto de resultados se reduce por la cantidad de filas que se omiten. Las filas omitidas por una cláusula OFFSET aún deben analizarse, por lo que puede ser ineficiente utilizar un valor OFFSET grande.

Notas de uso

Tenga en cuenta el siguiente comportamiento esperado con las cláusulas ORDER BY:

- Los valores NULL son considerados "superiores" a todos los otros valores. Con el orden ascendente predeterminado, los valores NULL se ordenan al final. Para cambiar este comportamiento, utilice la opción NULLS FIRST.
- Cuando una consulta no contiene una cláusula ORDER BY, el sistema devuelve conjuntos de resultados sin un orden predecible de las filas. Si se ejecuta la misma consulta dos veces, puede devolver el conjunto de resultados en un orden diferente.
- Las opciones LIMIT y OFFSET pueden utilizarse sin una cláusula ORDER BY; no obstante, para devolver un conjunto consistente de filas, use estas opciones junto con ORDER BY.
- En cualquier sistema paralelo como Amazon Redshift, cuando una cláusula ORDER BY no produce un solo orden, el orden de las filas no es determinístico Es decir, si la expresión ORDER BY produce valores duplicados, el orden devuelto de esas filas puede variar de otros sistemas o de una ejecución de Amazon Redshift a la siguiente.
- Amazon Redshift no admite literales de cadena en cláusulas ORDER BY.

### Ejemplos con ORDER BY

Devuelva todas las 11 filas de la tabla CATEGORY, ordenadas por la segunda columna, CATGROUP. Para los resultados que tienen el mismo valor CATGROUP, ordene los valores de la columna CATDESC por la longitud de la cadena de caracteres. Ordene, a continuación, por columna CATID y CATNAME.

```
select * from category order by 2, length(catdesc), 1, 3;

catid | catgroup |  catname  |                 catdesc
------+----------+-----------+--------------------------------------
10    | Concerts | Jazz      | All jazz singers and bands
9     | Concerts | Pop       | All rock and pop music concerts
11    | Concerts | Classical | All symphony, concerto, and choir conce
6     | Shows    | Musicals  | Musical theatre
7     | Shows    | Plays     | All non-musical theatre
8     | Shows    | Opera     | All opera and light opera
5     | Sports   | MLS       | Major League Soccer
1     | Sports   | MLB       | Major League Baseball
2     | Sports   | NHL       | National Hockey League
3     | Sports   | NFL       | National Football League
4     | Sports   | NBA       | National Basketball Association
(11 rows)
```

Devuelva las columnas seleccionadas de la tabla SALES, ordenadas por los valores QTYSOLD más altos. Limite el resultado a las primeras 10 filas:

```
select salesid, qtysold, pricepaid, commission, saletime from sales
order by qtysold, pricepaid, commission, salesid, saletime desc
limit 10;

salesid | qtysold | pricepaid | commission |      saletime
--------+---------+-----------+------------+---------------------
15401   |       8 |    272.00 |      40.80 | 2008-03-18 06:54:56
61683   |       8 |    296.00 |      44.40 | 2008-11-26 04:00:23
90528   |       8 |    328.00 |      49.20 | 2008-06-11 02:38:09
74549   |       8 |    336.00 |      50.40 | 2008-01-19 12:01:21
130232  |       8 |    352.00 |      52.80 | 2008-05-02 05:52:31
55243   |       8 |    384.00 |      57.60 | 2008-07-12 02:19:53
16004   |       8 |    440.00 |      66.00 | 2008-11-04 07:22:31
489     |       8 |    496.00 |      74.40 | 2008-08-03 05:48:55
4197    |       8 |    512.00 |      76.80 | 2008-03-23 11:35:33
16929   |       8 |    568.00 |      85.20 | 2008-12-19 02:59:33
(10 rows)
```

Devuelve una lista de columnas y ninguna fila a través de la sintaxis LIMIT 0:

```
select * from venue limit 0;
venueid | venuename | venuecity | venuestate | venueseats
---------+-----------+-----------+------------+------------
(0 rows)
```

## Cláusula CONNECT BY

La cláusula CONNECT BY especifica la relación entre las filas de una jerarquía. Puede utilizar CONNECT BY para seleccionar filas en orden jerárquico mediante la combinación de la tabla consigo misma y el procesamiento los datos jerárquicos. Por ejemplo, puede utilizarla para recorrer recursivamente un organigrama y enumerar los datos.

Las consultas jerárquicas se procesan en el siguiente orden:

1. Si la cláusula FROM tiene una combinación, se procesa primero.
2. Se evalúa la cláusula CONNECT BY.
3. Se evalúa la cláusula WHERE.

### Sintaxis

```
[START WITH start_with_conditions]
CONNECT BY connect_by_conditions
```

> **Note**
>
> Aunque START y CONNECT no son palabras reservadas, utilice identificadores delimitados (comillas dobles) o AS si utiliza START y CONNECT como alias de tabla en su consulta para evitar errores en tiempo de ejecución.

```
SELECT COUNT(*)
FROM Employee "start"
CONNECT BY PRIOR id = manager_id
START WITH name = 'John'
```

```
SELECT COUNT(*)
FROM Employee AS start
CONNECT BY PRIOR id = manager_id
START WITH name = 'John'
```

### Parámetros

start_with_conditions

Condiciones que especifican las filas raíz de la jerarquía

connect_by_conditions

Condiciones que especifican la relación entre las filas principales y las secundarias de la jerarquía. Se debe calificar al menos una condición con el operador unario que se usa para hacer referencia a la fila principal.

```
PRIOR column = expression
-- or
expression > PRIOR column
```

### Operadores

Puede utilizar los siguientes operadores en una consulta CONNECT BY.

LEVEL

Pseudocolumna que devuelve el nivel de fila actual en la jerarquía. Devuelve 1 para la fila raíz, 2 para el elemento secundario de la fila raíz, etc.

PRIOR

Operador unario que evalúa la expresión correspondiente a la fila principal de la fila actual de la jerarquía.

### Ejemplos

En el siguiente ejemplo se muestra una consulta CONNECT BY que devuelve la cantidad de empleados subordinados directos o indirectos de John, sin una profundidad superior a cuatro niveles.

```
SELECT id, name, manager_id
FROM employee
WHERE LEVEL < 4
START WITH name = 'John'
CONNECT BY PRIOR id = manager_id;
```

A continuación, se muestra el resultado de la consulta.

```
id      name      manager_id
------+----------+--------------
  101     John        100
  102     Jorge       101
  103     Kwaku       101
  110     Liu         101
  201     Sofía       102
  106     Mateo       102
  110     Nikki       103
  104     Paulo       103
  105     Richard     103
  120     Saanvi      104
  200     Shirley     104
  205     Zhang       104
```

Definición de la tabla para este ejemplo:

```
CREATE TABLE employee (
   id INT,
   name VARCHAR(20),
   manager_id INT
   );
```

A continuación, se presentan las filas insertadas en la tabla.

```
INSERT INTO employee(id, name, manager_id)  VALUES
(100, 'Carlos', null),
(101, 'John', 100),
(102, 'Jorge', 101),
(103, 'Kwaku', 101),
(110, 'Liu', 101),
(106, 'Mateo', 102),
(110, 'Nikki', 103),
(104, 'Paulo', 103),
(105, 'Richard', 103),
(120, 'Saanvi', 104),
(200, 'Shirley', 104),
(201, 'Sofía', 102),
(205, 'Zhang', 104);
```

A continuación, se muestra un gráfico de la organización del departamento de John.

## Ejemplos de subconsultas

En los siguientes ejemplos se muestran diferentes maneras en que las subconsultas encajan en las consultas SELECT. Para ver otro ejemplo del uso de las subconsultas, consulte Ejemplos de JOIN.

Subconsulta de la lista SELECT

En el siguiente ejemplo, se observa una subconsulta en la lista SELECT. Esta subconsulta es escalar: devuelve solamente una columna y un valor, que se repite en el resultado de cada fila que se devuelve desde la consulta externa. La consulta compara el valor Q1SALES que la subconsulta computa con valores de venta de otros dos trimestres (2 y 3) en 2008, como la consulta externa lo define.

```
select qtr, sum(pricepaid) as qtrsales,
(select sum(pricepaid)
from sales join date on sales.dateid=date.dateid
```

```
where qtr='1' and year=2008) as q1sales
from sales join date on sales.dateid=date.dateid
where qtr in('2','3') and year=2008
group by qtr
order by qtr;

qtr  |  qtrsales   |   q1sales
-------+-------------+-------------
2     | 30560050.00 | 24742065.00
3     | 31170237.00 | 24742065.00
(2 rows)
```

Subconsulta de la cláusula WHERE

En el siguiente ejemplo, se observa una subconsulta de tabla en la cláusula WHERE. Esta subconsulta produce varias filas. En este caso, las filas contienen solo una columna, pero las subconsultas de la tabla pueden contener varias columnas y filas, como cualquier otra tabla.

La consulta busca los principales 10 vendedores en términos de cantidad máxima de tickets vendidos. La lista de los 10 principales está limitada por la subconsulta, que elimina usuarios que viven en ciudades donde hay lugares de venta de tickets. Esta consulta puede escribirse en diferentes maneras; por ejemplo, se puede volver a escribir la subconsulta como una combinación dentro de la consulta principal.

```
select firstname, lastname, city, max(qtysold) as maxsold
from users join sales on users.userid=sales.sellerid
where users.city not in(select venuecity from venue)
group by firstname, lastname, city
order by maxsold desc, city desc
limit 10;

firstname | lastname  |      city      | maxsold
-----------+-----------+----------------+---------
Noah       | Guerrero | Worcester      |       8
Isadora    | Moss     | Winooski       |       8
Kieran     | Harrison | Westminster    |       8
Heidi      | Davis    | Warwick        |       8
Sara       | Anthony  | Waco           |       8
Bree       | Buck     | Valdez         |       8
Evangeline | Sampson  | Trenton        |       8
Kendall    | Keith    | Stillwater     |       8
Bertha     | Bishop   | Stevens Point  |       8
Patricia   | Anderson | South Portland |       8
```

```
(10 rows)
```

Subconsultas de la cláusula WITH

Consulte Cláusula WITH.

## Subconsultas correlacionadas

En el siguiente ejemplo, se observa una subconsulta correlacionada en la cláusula WHERE; este tipo de subconsulta contiene una o varias correlaciones entre sus columnas y las columnas producidas por la consulta externa. En este caso, la correlación es `where s.listid=l.listid`. Para cada fila que la consulta externa produce, se ejecuta la subconsulta para calificar o descalificar la fila.

```
select salesid, listid, sum(pricepaid) from sales s
where qtysold=
(select max(numtickets) from listing l
where s.listid=l.listid)
group by 1,2
order by 1,2
limit 5;

salesid | listid |   sum
--------+--------+----------
 27     |     28 | 111.00
 81     |    103 | 181.00
 142    |    149 | 240.00
 146    |    152 | 231.00
 194    |    210 | 144.00
(5 rows)
```

Patrones de subconsultas correlacionadas que no se admiten

El planificador de consultas usa un método de reescritura de consulta denominado descorrelación de subconsulta para optimizar varios patrones de subconsultas correlacionadas para la ejecución en un entorno MPP. Algunos tipos de subconsultas correlacionadas siguen patrones para los cuales Amazon Redshift no puede anular la correlación y que no admite. Las consultas que contienen las siguientes referencias de correlación devuelven errores:

- Referencias de correlación que omiten un bloque de consulta, también conocidas como "referencias de correlación con nivel omitido". Por ejemplo, en la siguiente consulta, el bloque que

contiene la referencia de correlación y el bloque omitido están conectados por un predicado NOT EXISTS:

```
select event.eventname from event
where not exists
(select * from listing
where not exists
(select * from sales where event.eventid=sales.eventid));
```

En este caso, el bloque omitido es la subconsulta que se ejecuta contra la tabla LISTING. La referencia de correlación correlaciona las tablas EVENT y SALES.

- Referencias de correlación de una subconsulta que es parte de una cláusula ON en una consulta externa:

```
select * from category
left join event
on category.catid=event.catid and eventid =
(select max(eventid) from sales where sales.eventid=event.eventid);
```

La cláusula ON contiene una referencia de correlación de SALES en la subconsulta a EVENT en la consulta externa.

- La correlación que distingue valores nulos referencia una tabla de sistema de Amazon Redshift. Por ejemplo:

```
select attrelid
from stv_locks sl, pg_attribute
where sl.table_id=pg_attribute.attrelid and 1 not in
(select 1 from pg_opclass where sl.lock_owner = opcowner);
```

- Referencias de correlación de una subconsulta que contiene una función de ventana.

```
select listid, qtysold
from sales s
where qtysold not in
(select sum(numtickets) over() from listing l where s.listid=l.listid);
```

- Referencias en una columna GROUP BY a los resultados de una subconsulta correlacionada. Por ejemplo:

```
select listing.listid,
(select count (sales.listid) from sales where sales.listid=listing.listid) as list
from listing
group by list, listing.listid;
```

- Referencias de correlación de una subconsulta con una función agregada y una cláusula GROUP BY, conectadas a la consulta externa por un predicado IN. (Esta restricción no se aplica a las funciones agregadas MIN y MAX). Por ejemplo:

```
select * from listing where listid in
(select sum(qtysold)
from sales
where numtickets>4
group by salesid);
```

# SELECT INTO

Selecciona las filas definidas por una consulta y las inserta en una nueva tabla. Puede especificar si va a crear una tabla temporal o persistente.

## Sintaxis

```
[ WITH with_subquery [, ...] ]
SELECT
[ TOP number | [ ALL | DISTINCT ]
* | expression [ AS output_name ] [, ...] ]
[ EXCLUDE column_list ]
INTO [ TEMPORARY | TEMP ] [ TABLE ] new_table
[ FROM table_reference [, ...] ]
[ WHERE condition ]
[ [ START WITH expression ] CONNECT BY expression ]
[ GROUP BY ALL | expression [, ...] ]
[ HAVING condition ]
[ QUALIFY condition ]
[ { UNION | ALL | INTERSECT | EXCEPT | MINUS } query ]
[ ORDER BY expression [ ASC | DESC ] ]
[ LIMIT { number | ALL } ]
[ OFFSET start ]
```

Para obtener información acerca de los parámetros de este comando, consulte SELECT.

### Ejemplos

Seleccione todas las filas de la tabla EVENT y cree la tabla NEWEVENT:

```
select * into newevent from event;
```

Seleccione el resultado de una consulta agregada en una tabla temporal denominada PROFITS:

```
select username, lastname, sum(pricepaid-commission) as profit
into temp table profits
from sales, users
where sales.sellerid=users.userid
group by 1, 2
order by 3 desc;
```

## SET

Establece el valor de un parámetro de configuración de un servidor. Utilice el comando SET para invalidar un valor solo mientras dure la sesión o la transacción actual.

Utilice el comando RESET para devolver un parámetro a su valor predeterminado.

Puede cambiar los parámetros de configuración del servidor de varias maneras. Para obtener más información, consulte Modificación de la configuración del servidor.

### Sintaxis

```
SET { [ SESSION | LOCAL ]
{ SEED | parameter_name } { TO | = }
{ value | 'value' | DEFAULT } |
SEED TO value }
```

La siguiente instrucción establece el valor de una variable de contexto de sesión.

```
SET { [ SESSION | LOCAL ]
variable_name { TO | = }
{ value | 'value'  }
```

## Parámetros

SESSION

Especifica que el ajuste es válido para la sesión actual. Valor predeterminado.

variable_name

Especifica el nombre del conjunto de variables de contexto de la sesión.

La convención de nomenclatura es un nombre de dos partes separado por un punto, por ejemplo identificador.identificador. Solo se permite un punto de separador. Utilice un identificador que sigue las reglas de identificador estándar de Amazon Redshift Para obtener más información, consulte Nombres e identificadores. No se permiten identificadores delimitados.

LOCAL

Especifica que el ajuste es válido para la transacción actual.

SEED TO value (valor)

Establece el uso de un inicio interno de la función RANDOM para la generación aleatoria de números.

SET SEED toma un valor numérico comprendido entre 0 y 1, y lo multiplica por ($2^{31}$-1) para utilizarlo con la función Función RANDOM. Si utiliza SET SEED antes de realizar varias ejecuciones RANDOM, RANDOM genera números en una secuencia predecible.

parameter_name

Nombre del parámetro que se debe establecer. Para obtener información acerca de los parámetros, consulte Modificación de la configuración del servidor.

value

Nuevo valor del parámetro. Utilice comillas simples para establecer el valor en una cadena específica. Si utiliza SET SEED, este parámetro contiene el valor SEED.

DEFAULT

Establece el parámetro al valor predeterminado.

## Ejemplos

Cambio de un parámetro en la sesión actual

En el siguiente ejemplo, se establece el estilo de fecha:

```
set datestyle to 'SQL,DMY';
```

Establecimiento de un grupo de consultas para la administración de cargas de trabajo

Si los grupos de consultas están enumerados en la definición de una cola como parte de la configuración WLM del clúster, puede establecer el parámetro QUERY_GROUP con el nombre de un grupo de consultas enumerado. Las consultas posteriores se asignan a la cola de la consulta asociada. El ajuste QUERY_GROUP permanece en vigor durante la sesión o hasta que se aplica un comando RESET QUERY_GROUP.

Este ejemplo ejecuta dos consultas como parte de la "prioridad" del grupo de consultas y, luego, restablece el grupo de consultas.

```
set query_group to 'priority';
select tbl, count(*)from stv_blocklist;
select query, elapsed, substring from svl_qlog order by query desc limit 5;
reset query_group;
```

Para obtener más información, consulte Administración de la carga de trabajo.

Cambio del espacio de nombres de identidad predeterminado para la sesión

Un usuario de la base de datos puede establecer `default_identity_namespace`. En este ejemplo, se muestra cómo usar `SET SESSION` para invalidar la configuración durante la sesión actual y, a continuación, mostrar el nuevo valor del proveedor de identidades. Esto se utiliza con más frecuencia cuando utiliza un proveedor de identidades con Redshift e IAM Identity Center. Para obtener más información sobre cómo usar un proveedor de identidades con Redshift, consulte Conectar Redshift con IAM Identity Center para ofrecer a los usuarios una experiencia de inicio de sesión único.

```
SET SESSION default_identity_namespace = 'MYCO';

SHOW default_identity_namespace;
```

Tras ejecutar el comando, puede ejecutar una instrucción GRANT o CREATE como la siguiente:

```
GRANT SELECT ON TABLE mytable TO alice;
```

```
GRANT UPDATE ON TABLE mytable TO salesrole;

CREATE USER bob password 'md50c983d1a624280812631c5389e60d48c';
```

En este caso, el efecto de establecer el espacio de nombres de identidad predeterminado equivale a anteponer cada identidad con el espacio de nombres. En este ejemplo, `alice` se sustituye por `MYCO:alice`. Para obtener más información acerca de los ajustes relacionados con la configuración de Redshift con IAM Identity Center, consulte ALTER SYSTEM y MODIFICAR PROVEEDOR DE IDENTIDADES.

Establecimiento de una etiqueta para un grupo de consultas

El parámetro QUERY_GROUP define una etiqueta para una o más consultas que se ejecutan en la misma sesión después de un comando SET. A su vez, se registra esta etiqueta cuando se ejecutan las consultas y se puede utilizar para limitar los resultados que devuelven las tablas de sistema STL_QUERY y STV_INFLIGHT, además de la vista SVL_QLOG.

```
show query_group;
query_group
-------------
unset
(1 row)

set query_group to '6 p.m.';


show query_group;
query_group
-------------
6 p.m.
(1 row)

select * from sales where salesid=500;
salesid | listid | sellerid | buyerid | eventid | dateid | ...
---------+--------+----------+---------+---------+--------+-----
500 |    504 |     3858 |    2123 |    5871 |   2052 | ...
(1 row)

reset query_group;

select query, trim(label) querygroup, pid, trim(querytxt) sql
from stl_query
```

```
where label ='6 p.m.';
query | querygroup |  pid  |                  sql
-------+------------+-------+----------------------------------------
57 | 6 p.m.     | 30711 | select * from sales where salesid=500;
(1 row)
```

Las etiquetas del grupo de consultas son un mecanismo útil para aislar consultas individuales o grupos de consultas que se ejecutan como parte de scripts. No es necesario identificar y rastrear las consultas por sus ID; puede rastrearlas por sus etiquetas.

Establecimiento de un valor de inicio para la generación aleatoria de números

En el siguiente ejemplo, se utiliza la opción SEED con SET para que la función RANDOM genere números en una secuencia predecible.

Primero, se devuelven tres valores enteros RANDOM sin establecer antes el valor SEED:

```
select cast (random() * 100 as int);
int4
------
6
(1 row)

select cast (random() * 100 as int);
int4
------
68
(1 row)

select cast (random() * 100 as int);
int4
------
56
(1 row)
```

Ahora, establezca el valor SEED en `.25` y devuelva tres números RANDOM más:

```
set seed to .25;

select cast (random() * 100 as int);
int4
------
```

```
21
(1 row)

select cast (random() * 100 as int);
int4
------
79
(1 row)

select cast (random() * 100 as int);
int4
------
12
(1 row)
```

Finalmente, restablezca el valor SEED a `.25` y verifique que RANDOM devuelva los mismos resultados que en las tres ejecuciones anteriores:

```
set seed to .25;

select cast (random() * 100 as int);
int4
------
21
(1 row)

select cast (random() * 100 as int);
int4
------
79
(1 row)

select cast (random() * 100 as int);
int4
------
12
(1 row)
```

En el siguiente ejemplo, se establece una variable de contexto personalizada.

```
SET app_context.user_id TO 123;
SET app_context.user_id TO 'sample_variable_value';
```

# SET SESSION AUTHORIZATION

Establece el nombre de usuario para la sesión actual.

Puede usar el comando SET SESSION AUTHORIZATION, por ejemplo, para probar el acceso a la base de datos al ejecutar temporalmente una sesión o transacción como usuario sin privilegios. Debe ser un superusuario de base de datos para ejecutar este comando.

## Sintaxis

```
SET [ LOCAL ] SESSION AUTHORIZATION { user_name | DEFAULT }
```

## Parámetros

LOCAL

Especifica que el ajuste es válido para la transacción actual. Omitir este parámetro especifica que el ajuste es válido para la sesión actual.

user_name

Nombre del usuario que se debe establecer. El nombre de usuario puede escribirse como identificador o como literal de cadena.

DEFAULT

Establece el nombre de usuario de la sesión al valor predeterminado.

## Ejemplos

En el siguiente ejemplo, se establece el nombre de usuario para la sesión actual en `dwuser`:

```
SET SESSION AUTHORIZATION 'dwuser';
```

En el siguiente ejemplo, se establece el nombre de usuario para la transacción actual en `dwuser`:

```
SET LOCAL SESSION AUTHORIZATION 'dwuser';
```

Este siguiente ejemplo establece el nombre de usuario para la sesión actual en el nombre de usuario predeterminado:

```
SET SESSION AUTHORIZATION DEFAULT;
```

# SET SESSION CHARACTERISTICS

Este comando no está disponible.

# SHOW

Muestra el valor actual de un parámetro de configuración de un servidor. Este valor puede ser específico de la sesión actual si hay un comando SET en vigor. Para obtener una lista de parámetros de configuración, consulte Referencia de la configuración.

## Sintaxis

```
SHOW { parameter_name | ALL }
```

La siguiente instrucción muestra el valor actual de una variable de contexto de sesión. Si la variable no existe, Amazon Redshift emite un error.

```
SHOW variable_name
```

## Parámetros

parameter_name

Muestra el valor actual del parámetro especificado.

ALL

Muestra los valores actuales de todos los parámetros.

variable_name

Muestra el valor actual de la variable especificada.

## Ejemplos

En el siguiente ejemplo, se muestra el valor para el parámetro query_group:

```
show query_group;
```

```
query_group

unset
(1 row)
```

En el siguiente ejemplo, se muestra una lista de todos los parámetros y sus valores:

```
show all;
name        |   setting
--------------------+--------------
datestyle           | ISO, MDY
extra_float_digits | 0
query_group         | unset
search_path         | $user,public
statement_timeout   | 0
```

En el siguiente ejemplo se muestra el valor actual de la variable especificada.

```
SHOW app_context.user_id;
```

# SHOW COLUMNS

Muestra una lista de columnas de una tabla, junto con algunos atributos de columna.

Cada fila de salida consta de una lista separada por comas con el nombre de la base de datos, el nombre del esquema, el nombre de la tabla, el nombre de la columna, la posición ordinal, la columna predeterminada, indicación de si se puede anular, el tipo de datos, la longitud máxima de los caracteres, la precisión numérica y las observaciones. Para obtener más información sobre estos atributos, consulte SVV_ALL_COLUMNS.

Si el comando SHOW COLUMNS diera como resultado más de 10 000 columnas, se devolvería un error.

## Permisos necesarios

Para ver una columna de una tabla de Amazon Redshift, el usuario actual debe cumplir uno de los siguientes criterios:

- Ser un superusuario.
- Ser el propietario de la tabla.

- Tener concedido el privilegio USAGE en el esquema principal y el privilegio SELECT en la tabla o el privilegio SELECT en la columna.

## Sintaxis

```
SHOW COLUMNS FROM TABLE database_name.schema_name.table_name [LIKE 'filter_pattern']
 [LIMIT row_limit ]
```

## Parámetros

database_name

Nombre de la base de datos que contiene las tablas que se enumerarán.

Para mostrar tablas en un AWS Glue Data Catalog, especifique (`awsdatacatalog`) como nombre de base de datos y asegúrese de que la configuración del sistema `data_catalog_auto_mount` está establecida a `true`. Para obtener más información, consulte ALTER SYSTEM.

schema_name

Nombre del esquema que contiene las tablas que se enumerarán.

Para mostrar las tablas de AWS Glue Data Catalog, indique el nombre de la base de datos de AWS Glue como nombre del esquema.

table_name

Nombre de la tabla que contiene las columnas que se enumerarán.

filter_pattern

Expresión de caracteres UTF-8 válida con un patrón para hacer coincidir los nombres de tabla. La opción LIKE realiza una coincidencia que distingue entre mayúsculas y minúsculas y admite los siguientes metacaracteres de coincidencia de patrones:

| Metacarácter | Descripción |
|---|---|
| % | Coincide con cualquier secuencia de cero o más caracteres. |
| _ | Coincide con cualquier carácter. |

Si filter_pattern no contiene metacaracteres, solo representa la propia cadena; en ese caso, LIKE actúa igual que el operador de igualdad.

row_limit

Número máximo de filas que se devolverán. El valor de row_limit está entre 0 y 10 000.

## Ejemplos

En el siguiente ejemplo se muestran las columnas de la base de datos de Amazon Redshift denominada `sample_data_dev` que se encuentran en el esquema `tickit` y la tabla `event`.

```
SHOW COLUMNS FROM TABLE sample_data_dev.tickit.event;

 database_name     | schema_name | table_name | column_name | ordinal_position |
 column_default | is_nullable | data_type                    | character_maximum_length
 | numeric_precision | numeric_scale | remarks
-------------------+-------------+------------+-------------+------------------
+----------------+-------------+-----------------------------
+--------------------------+-------------------+-----------------------
 sample_data_dev   | tickit      | event      | eventid     |                1 | NULL
         | NO          | integer                      | NULL                     |
           32 | 0             | NULL
 sample_data_dev   | tickit      | event      | eventid     |                2 | NULL
         | NO          | smallint                     | NULL                     |
           16 | 0             | NULL
 sample_data_dev   | tickit      | event      | eventid     |                3 | NULL
         | NO          | smallint                     | NULL                     |
           16 | 0             | NULL
 sample_data_dev   | tickit      | event      | eventid     |                4 | NULL
         | NO          | smallint                     | NULL                     |
           16 | 0             | NULL
 sample_data_dev   | tickit      | event      | eventid     |                5 | NULL
         | YES         | character varying            | 200                      |
         NULL | NULL          | NULL
 sample_data_dev   | tickit      | event      | eventid     |                6 | NULL
         | YES         | timestamp without time zo... | NULL                     |
         NULL | NULL          | NULL
```

En el siguiente ejemplo se muestran las columnas de la base de datos de AWS Glue Data Catalog denominada `awsdatacatalog` que se encuentran en el esquema `batman` y la tabla `nation`. La salida está limitada a 2 filas.

```
SHOW COLUMNS FROM TABLE awsdatacatalog.batman.nation LIMIT 2;

 database_name  | schema_name | table_name | column_name | ordinal_position
 | column_default | is_nullable | data_type | character_maximum_length |
 numeric_precision | remarks
----------------+-------------+------------+-------------+-----------------
+----------------+-------------+-----------+-------------------------
+-------------------+---------
 awsdatacatalog | batman      | nation     | n_nationkey |               1 |
      |             | integer   |                         |                 |
 awsdatacatalog | batman      | nation     | n_name      |               2 |
      |             | character |                         |                 |
```

## SHOW EXTERNAL TABLE

Muestra la definición de una tabla externa, incluidos los atributos de tabla y los atributos de columna. Puede utilizar la salida de la instrucción SHOW EXTERNAL TABLE para recrear la tabla.

Para obtener más información acerca de la creación de tablas externas, consulte CREATE EXTERNAL TABLE.

### Sintaxis

```
SHOW EXTERNAL TABLE [external_database].external_schema.table_name [ PARTITION ]
```

### Parámetros

external_database

Se trata del nombre de la base de datos externa asociada. Este parámetro es opcional.

external_schema

Se trata del nombre del esquema externo asociado.

table_name

Se trata del nombre de la tabla que se mostrará.

PARTITION

Muestra instrucciones ALTER TABLE para agregar particiones a la definición de la tabla.

## Ejemplos

Los siguientes ejemplos se basan en una tabla externa definida de la siguiente manera:

```
CREATE EXTERNAL TABLE my_schema.alldatatypes_parquet_test_partitioned (
     csmallint smallint,
     cint int,
     cbigint bigint,
     cfloat float4,
     cdouble float8,
     cchar char(10),
     cvarchar varchar(255),
     cdecimal_small decimal(18,9),
     cdecimal_big decimal(30,15),
     ctimestamp TIMESTAMP,
     cboolean boolean,
     cstring varchar(16383)
)
PARTITIONED BY (cdate date, ctime TIMESTAMP)
STORED AS PARQUET
LOCATION 's3://amzn-s3-demo-bucket/alldatatypes_parquet_partitioned';
```

A continuación, se presenta un ejemplo del comando SHOW EXTERNAL TABLE y la salida para la tabla `my_schema.alldatatypes_parquet_test_partitioned`.

```
SHOW EXTERNAL TABLE my_schema.alldatatypes_parquet_test_partitioned;
```

```
"CREATE EXTERNAL TABLE my_schema.alldatatypes_parquet_test_partitioned (
    csmallint smallint,
    cint int,
    cbigint bigint,
    cfloat float4,
    cdouble float8,
    cchar char(10),
    cvarchar varchar(255),
    cdecimal_small decimal(18,9),
    cdecimal_big decimal(30,15),
```

```
    ctimestamp timestamp,
    cboolean boolean,
    cstring varchar(16383)
)
PARTITIONED BY (cdate date, ctime timestamp)
ROW FORMAT SERDE 'org.apache.hadoop.hive.ql.io.parquet.serde.ParquetHiveSerDe'
STORED AS INPUTFORMAT 'org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat'
OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat'
LOCATION 's3://amzn-s3-demo-bucket/alldatatypes_parquet_partitioned';"
```

A continuación, se presenta un ejemplo del comando SHOW EXTERNAL TABLE y la salida para la misma tabla, pero con la base de datos especificada también en el parámetro.

```
SHOW EXTERNAL TABLE my_database.my_schema.alldatatypes_parquet_test_partitioned;
```

```
"CREATE EXTERNAL TABLE my_database.my_schema.alldatatypes_parquet_test_partitioned (
    csmallint smallint,
    cint int,
    cbigint bigint,
    cfloat float4,
    cdouble float8,
    cchar char(10),
    cvarchar varchar(255),
    cdecimal_small decimal(18,9),
    cdecimal_big decimal(30,15),
    ctimestamp timestamp,
    cboolean boolean,
    cstring varchar(16383)
)
PARTITIONED BY (cdate date, ctime timestamp)
ROW FORMAT SERDE 'org.apache.hadoop.hive.ql.io.parquet.serde.ParquetHiveSerDe'
STORED AS INPUTFORMAT 'org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat'
OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat'
LOCATION 's3://amzn-s3-demo-bucket/alldatatypes_parquet_partitioned';"
```

A continuación, se presenta un ejemplo del comando SHOW EXTERNAL TABLE y la salida cuando se utiliza el parámetro `PARTITION`. La salida contiene instrucciones ALTER TABLE para agregar particiones a la definición de la tabla.

```
SHOW EXTERNAL TABLE my_schema.alldatatypes_parquet_test_partitioned PARTITION;
```

```
"CREATE EXTERNAL TABLE my_schema.alldatatypes_parquet_test_partitioned (
    csmallint smallint,
    cint int,
    cbigint bigint,
    cfloat float4,
    cdouble float8,
    cchar char(10),
    cvarchar varchar(255),
    cdecimal_small decimal(18,9),
    cdecimal_big decimal(30,15),
    ctimestamp timestamp,
    cboolean boolean,
    cstring varchar(16383)
)
PARTITIONED BY (cdate date)
ROW FORMAT SERDE 'org.apache.hadoop.hive.ql.io.parquet.serde.ParquetHiveSerDe'
STORED AS INPUTFORMAT 'org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat'
OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat'
LOCATION 's3://amzn-s3-demo-bucket/alldatatypes_parquet_partitioned';
ALTER TABLE my_schema.alldatatypes_parquet_test_partitioned ADD IF NOT
 EXISTS PARTITION (cdate='2021-01-01') LOCATION 's3://amzn-s3-demo-bucket/
alldatatypes_parquet_partitioned2/cdate=2021-01-01';
ALTER TABLE my_schema.alldatatypes_parquet_test_partitioned ADD IF NOT
 EXISTS PARTITION (cdate='2021-01-02') LOCATION 's3://amzn-s3-demo-bucket/
alldatatypes_parquet_partitioned2/cdate=2021-01-02';"
```

# SHOW DATABASES

Muestra las bases de datos de un catálogo de datos o un almacenamiento de datos de Amazon Redshift. SHOW DATABASES enumera todas las bases de datos accesibles, como las que se encuentran dentro del almacenamiento de datos, las bases de datos de AWS Glue Data Catalog (awsdatacatalog), las bases de datos de uso compartido de datos y las bases de datos de Lake Formation.

## Permisos necesarios

Los usuarios pueden ver todas las bases de datos, excepto:

- En el caso de las bases de datos creadas a partir de un recurso compartido de datos con permisos para ser visibles, se debe conceder al usuario actual el permiso USAGE en la base de datos.

## Sintaxis

Para mostrar las bases de datos de un almacenamiento de datos de Amazon Redshift:

```
SHOW DATABASES
[ LIKE '<expression>' ]
[ LIMIT row_limit ]
```

Para mostrar las bases de datos de un catálogo de datos:

```
SHOW DATABASES FROM DATA CATALOG
[ ACCOUNT  '<id1>', '<id2>', ... ]
[ LIKE '<expression>' ]
[ IAM_ROLE default | 'SESSION' | 'arn:aws:iam::<account-id>:role/<role-name>' ]
[ LIMIT row_limit ]
```

## Parámetros

ACCOUNT '<id1>', '<id2>', ...

Las cuentas de AWS Glue Data Catalog desde las que se enumeran las bases de datos. La omisión de este parámetro indica que Amazon Redshift debe mostrar las bases de datos de la cuenta propietaria del clúster.

LIKE '<expresión>'

Filtra la lista de bases de datos para obtener las que coinciden con la que especifica. Este parámetro admite patrones que utilizan los caracteres comodín % (porcentaje) y _ (subrayado).

IAM_ROLE default | 'SESSION' | 'arn:aws:iam::<id-cuenta>:role/<nombre-rol>'

Si especifica un rol de IAM asociado al clúster al ejecutar el comando SHOW DATABASES, Amazon Redshift utilizará las credenciales del rol cuando ejecute consultas en la base de datos.

Especificar la palabra clave `default` significa utilizar el rol de IAM establecido como predeterminado y asociado al clúster.

Use `'SESSION'` si se conecta al clúster de Amazon Redshift mediante una identidad federada y acceda a las tablas desde la base de datos externa creada con el comando the section called “CREATE DATABASE”. Para ver un ejemplo de utilización de una identidad federada, consulte Uso de una identidad federada para administrar el acceso de Amazon Redshift a los recursos

locales y a las tablas externas de Amazon Redshift Spectrum, lo que explica cómo configurar la identidad federada.

Utilice el nombre de recurso de Amazon (ARN), de un rol de IAM que el clúster utiliza para la autenticación y la autorización. Como mínimo, el rol de IAM debe tener permiso para realizar una operación LIST en el bucket de Amazon S3 al que se accederá y una operación GET en los objetos de Amazon S3 que el bucket contiene. Para obtener más información sobre las bases de datos creadas a partir AWS Glue Data Catalog de los recursos compartidos de datos y utilizar IAM_ROLE, consulte Uso de los recursos compartidos de datos administrados por Lake Formation como consumidor.

A continuación se muestra la sintaxis de la cadena del parámetro IAM_ROLE para un único ARN.

```
IAM_ROLE 'arn:aws:iam::<aws-account-id>:role/<role-name>'
```

Puede encadenar roles para que el clúster pueda asumir otro rol de IAM, que posiblemente pertenezca a otra cuenta. Puede encadenar hasta 10 roles. Para obtener más información, consulte Encadenamiento de roles de IAM en Amazon Redshift Spectrum.

Para este rol de IAM; asocie una política de permisos de IAM similar a la siguiente.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AccessSecret",
            "Effect": "Allow",
            "Action": [
                "secretsmanager:GetResourcePolicy",
                "secretsmanager:GetSecretValue",
                "secretsmanager:DescribeSecret",
                "secretsmanager:ListSecretVersionIds"
            ],
            "Resource": "arn:aws:secretsmanager:us-west-2:123456789012:secret:my-
rds-secret-VNenFy"
        },
        {
            "Sid": "VisualEditor1",
            "Effect": "Allow",
            "Action": [
```

```
                "secretsmanager:GetRandomPassword",
                "secretsmanager:ListSecrets"
            ],
            "Resource": "*"
        }
    ]
}
```

Para obtener información sobre los pasos para crear un rol de IAM que se utilizará con la consulta federada, consulte Creación de un secreto y rol de IAM para utilizar consultas federadas.

> **Note**
>
> No incluya espacios en la lista de roles encadenados.

A continuación se muestra la sintaxis para encadenar tres roles.

```
IAM_ROLE 'arn:aws:iam::<aws-account-id>:role/<role-1-name>,arn:aws:iam::<aws-
account-id>:role/<role-2-name>,arn:aws:iam::<aws-account-id>:role/<role-3-name>'
```

LIMIT row_limit

Cláusula para LIMITAR la cantidad de filas devueltas Donde row_limit es el número máximo de filas que se devolverán. El valor de row_limit está entre 0 y 10 000.

## Ejemplos

En el siguiente ejemplo, se muestran todas las bases de datos del catálogo de datos del ID de cuenta 123456789012.

```
SHOW DATABASES FROM DATA CATALOG ACCOUNT '123456789012'

  catalog_id  | database_name |                        database_arn
 |     type     |                                          target_database
                           | location | parameters
--------------+---------------+-----------------------------------------------------------
+--------------
+-------------------------------------------------------------------------------------------
+----------+------------
```

```
 123456789012 |   database1   | arn:aws:glue:us-east-1:123456789012:database/database1
 | Data Catalog |
                            |          |
 123456789012 |   database2   | arn:aws:glue:us-east-1:123456789012:database/database2
 | Data Catalog | arn:aws:redshift:us-
east-1:123456789012:datashare:035c45ea-61ce-86f0-8b75-19ac6102c3b7/database2 |
 |
```

Los siguientes son ejemplos que demuestran cómo visualizar todas las bases de datos del catálogo de datos desde el ID de cuenta 123456789012 utilizando las credenciales de un rol de IAM.

```
SHOW DATABASES FROM DATA CATALOG ACCOUNT '123456789012' IAM_ROLE default;
```

```
SHOW DATABASES FROM DATA CATALOG ACCOUNT '123456789012' IAM_ROLE <iam-role-arn>;
```

En el siguiente ejemplo, se muestran todas las bases de datos del almacenamiento de datos de Amazon Redshift conectado.

```
SHOW DATABASES

database_name  | database_owner | database_type        | database_acl | parameters |
 database_isolation_level
---------------+----------------+----------------------+--------------+------------
+--------------------
awsdatacatalog | 1              | auto mounted catalog | NULL         | UNKNOWN    |
 UNKNOWN
dev            | 1              | local                | NULL         | NULL       |
 Snapshot Isolation
```

# SHOW GRANTS

Muestra las concesiones para un usuario, rol u objeto. El objeto puede ser una base de datos, un esquema, una tabla o una función. Al especificar un objeto, como una tabla o una función, debe calificarlo con una notación de dos o tres partes. Por ejemplo, `schema_name.table_name` o `database_name.schema_name.table_name`.

## Sintaxis

A continuación, se muestra la sintaxis para mostrar los permisos de un objeto. Tenga en cuenta que la segunda forma de especificar una función solo es válida para esquemas y bases de datos externos creados a partir de un recurso compartido de datos.

```
SHOW GRANTS ON
{
 DATABASE database_name |
 FUNCTION {database_name.schema_name.function_name | schema_name.function_name }
 ( [ [ argname ] argtype [, ...] ] ) |
 FUNCTION {database_name.schema_name.function_name | schema_name.function_name } |
 SCHEMA {database_name.schema_name | schema_name} |
 { TABLE {database_name.schema_name.table_name | schema_name.table_name} | table_name }
}
[FOR {username | ROLE role_name | PUBLIC}]
[LIMIT row_limit]
```

A continuación, se muestra la sintaxis para mostrar los permisos de un usuario o rol.

```
SHOW GRANTS FOR
{username | ROLE role_name}
[LIMIT row_limit]
```

## Parámetros

database_name

Nombre de la base de datos en la que se muestran las concesiones.

function_name

Nombre de la función en la que se muestran las concesiones.

schema_name

Nombre del esquema en el que se muestran las concesiones.

table_name

Nombre de la tabla en la que se muestran las concesiones.

FOR username

Indica que se muestran concesiones para un usuario.

FOR ROLE role_name

Indica que se muestran concesiones para un rol.

FOR PUBLIC

Indica que se muestran concesiones para PUBLIC.

row_limit

Número máximo de filas que se devolverán. El valor de row_limit está entre 0 y 10 000.

## Ejemplos

El siguiente ejemplo muestra todas las concesiones en una base de datos denominada dev.

```
SHOW GRANTS ON DATABASE dev;

 database_name | privilege_type | identity_id | identity_name | identity_type |
 admin_option | privilege_scope
---------------+----------------+-------------+---------------+---------------
+--------------+-----------------
 dev           | TRUNCATE       |         101 | alice         | user          | f
      | TABLES
 dev           | DROP           |         101 | alice         | user          | f
      | TABLES
 dev           | INSERT         |         101 | alice         | user          | f
      | TABLES
 dev           | ALTER          |         101 | alice         | user          | f
      | TABLES
 dev           | TEMP           |           0 | public        | public        | f
      | DATABASE
 dev           | DELETE         |         101 | alice         | user          | f
      | TABLES
 dev           | SELECT         |         101 | alice         | user          | f
      | TABLES
 dev           | UPDATE         |         101 | alice         | user          | f
      | TABLES
 dev           | REFERENCES     |         101 | alice         | user          | f
      | TABLES
(9 rows)
```

El siguiente comando muestra todas las concesiones en un esquema denominado demo.

```
SHOW GRANTS ON SCHEMA demo;

 schema_name | object_name | object_type | privilege_type | identity_id | identity_name
 | identity_type | admin_option | privilege_scope
-------------+-------------+-------------+----------------+-------------
+---------------+---------------+--------------+-----------------
 demo        | demo        | SCHEMA      | ALTER          |         101 | alice
 | user          | f             | SCHEMA
 demo        | demo        | SCHEMA      | DROP           |         101 | alice
 | user          | f             | SCHEMA
 demo        | demo        | SCHEMA      | USAGE          |         101 | alice
 | user          | f             | SCHEMA
 demo        | demo        | SCHEMA      | CREATE         |         101 | alice
 | user          | f             | SCHEMA
(4 rows)
```

El siguiente comando muestra todas las concesiones para un usuario denominado `alice`.

```
SHOW GRANTS FOR alice;

 database_name | schema_name | object_name | object_type | privilege_type | identity_id
 | identity_name | identity_type | privilege_scope
---------------+-------------+-------------+-------------+----------------
+-------------+---------------+---------------+-----------------
 dev           |             |             | DATABASE    | INSERT         |         101
 | alice         | user          | TABLES
 dev           |             |             | DATABASE    | SELECT         |         101
 | alice         | user          | TABLES
 dev           |             |             | DATABASE    | UPDATE         |         101
 | alice         | user          | TABLES
 dev           |             |             | DATABASE    | DELETE         |         101
 | alice         | user          | TABLES
 dev           |             |             | DATABASE    | REFERENCES     |         101
 | alice         | user          | TABLES
 dev           |             |             | DATABASE    | DROP           |         101
 | alice         | user          | TABLES
 dev           |             |             | DATABASE    | TRUNCATE       |         101
 | alice         | user          | TABLES
 dev           |             |             | DATABASE    | ALTER          |         101
 | alice         | user          | TABLES
 dev           | public      | t1          | TABLE       | INSERT         |         101
 | alice         | user          | TABLE
```

```
 dev           | public      | t1          | TABLE       | SELECT         |         101
 | alice         | user          | TABLE
 dev           | public      | t1          | TABLE       | UPDATE         |         101
 | alice         | user          | TABLE
 dev           | public      | t1          | TABLE       | DELETE         |         101
 | alice         | user          | TABLE
 dev           | public      | t1          | TABLE       | REFERENCES     |         101
 | alice         | user          | TABLE
 dev           | public      | t1          | TABLE       | DROP           |         101
 | alice         | user          | TABLE
 dev           | public      | t1          | TABLE       | TRUNCATE       |         101
 | alice         | user          | TABLE
 dev           | public      | t1          | TABLE       | ALTER          |         101
 | alice         | user          | TABLE
 dev           | demo        |             | SCHEMA      | USAGE          |         101
 | alice         | user          | SCHEMA
 dev           | demo        |             | SCHEMA      | CREATE         |         101
 | alice         | user          | SCHEMA
 dev           | demo        |             | SCHEMA      | DROP           |         101
 | alice         | user          | SCHEMA
 dev           | demo        |             | SCHEMA      | ALTER          |         101
 | alice         | user          | SCHEMA
(20 rows)
```

El siguiente comando muestra todas las concesiones en una tabla denominada `t3` para un usuario denominado `alice`. Tenga en cuenta que puede utilizar una notación de dos o tres partes para especificar el nombre de la tabla.

```
SHOW GRANTS ON TABLE demo_db.demo_schema.t3 FOR alice;
 schema_name | object_name | object_type | privilege_type | identity_id | identity_name
 | identity_type | admin_option | privilege_scope
-------------+-------------+-------------+----------------+-------------
+---------------+---------------+--------------+-----------------
 demo_schema | t3          | TABLE       | ALTER          |         100 | alice
 | user          | f             | TABLE
 demo_schema | t3          | TABLE       | TRUNCATE       |         100 | alice
 | user          | f             | TABLE
 demo_schema | t3          | TABLE       | DROP           |         100 | alice
 | user          | f             | TABLE
 demo_schema | t3          | TABLE       | SELECT         |         100 | alice
 | user          | f             | TABLE
```

```
 demo_schema | t3          | TABLE       | INSERT         |         100 | alice
 | user          | f            | TABLE
 demo_schema | t3          | TABLE       | UPDATE         |         100 | alice
 | user          | f            | TABLE
 demo_schema | t3          | TABLE       | DELETE         |         100 | alice
 | user          | f            | TABLE
 demo_schema | t3          | TABLE       | REFERENCES     |         100 | alice
 | user          | f            | TABLE
(8 rows)

SHOW GRANTS ON TABLE demo_schema.t3 FOR alice;
 schema_name | object_name | object_type | privilege_type | identity_id | identity_name
 | identity_type | admin_option | privilege_scope
-------------+-------------+-------------+----------------+-------------
+---------------+---------------+--------------+-----------------
 demo_schema | t3          | TABLE       | ALTER          |         100 | alice
 | user          | f            | TABLE
 demo_schema | t3          | TABLE       | TRUNCATE       |         100 | alice
 | user          | f            | TABLE
 demo_schema | t3          | TABLE       | DROP           |         100 | alice
 | user          | f            | TABLE
 demo_schema | t3          | TABLE       | SELECT         |         100 | alice
 | user          | f            | TABLE
 demo_schema | t3          | TABLE       | INSERT         |         100 | alice
 | user          | f            | TABLE
 demo_schema | t3          | TABLE       | UPDATE         |         100 | alice
 | user          | f            | TABLE
 demo_schema | t3          | TABLE       | DELETE         |         100 | alice
 | user          | f            | TABLE
 demo_schema | t3          | TABLE       | REFERENCES     |         100 | alice
 | user          | f            | TABLE
(8 rows)
```

El siguiente ejemplo muestra todas las concesiones en una tabla denominada `t4`. Tenga en cuenta las diferentes formas en que puede especificar el nombre de la tabla.

```
SHOW GRANTS ON t4;
 schema_name | object_name | object_type | privilege_type | identity_id | identity_name
 | identity_type | admin_option | privilege_scope
-------------+-------------+-------------+----------------+-------------
+---------------+---------------+--------------+-----------------
 public      | t4          | TABLE       | ALTER          |         100 | alice
 | user          | f            | TABLE
```

```
 public      | t4          | TABLE       | TRUNCATE       |         100 | alice
 | user          | f            | TABLE
 public      | t4          | TABLE       | DROP           |         100 | alice
 | user          | f            | TABLE
 public      | t4          | TABLE       | SELECT         |         100 | alice
 | user          | f            | TABLE
 public      | t4          | TABLE       | INSERT         |         100 | alice
 | user          | f            | TABLE
 public      | t4          | TABLE       | UPDATE         |         100 | alice
 | user          | f            | TABLE
 public      | t4          | TABLE       | DELETE         |         100 | alice
 | user          | f            | TABLE
 public      | t4          | TABLE       | REFERENCES     |         100 | alice
 | user          | f            | TABLE
(8 rows)

SHOW GRANTS ON TABLE public.t4;
 schema_name | object_name | object_type | privilege_type | identity_id | identity_name
 | identity_type | admin_option | privilege_scope
-------------+-------------+-------------+----------------+-------------
+---------------+---------------+--------------+-----------------
 public      | t4          | TABLE       | ALTER          |         100 | alice
 | user          | f            | TABLE
 public      | t4          | TABLE       | TRUNCATE       |         100 | alice
 | user          | f            | TABLE
 public      | t4          | TABLE       | DROP           |         100 | alice
 | user          | f            | TABLE
 public      | t4          | TABLE       | SELECT         |         100 | alice
 | user          | f            | TABLE
 public      | t4          | TABLE       | INSERT         |         100 | alice
 | user          | f            | TABLE
 public      | t4          | TABLE       | UPDATE         |         100 | alice
 | user          | f            | TABLE
 public      | t4          | TABLE       | DELETE         |         100 | alice
 | user          | f            | TABLE
 public      | t4          | TABLE       | REFERENCES     |         100 | alice
 | user          | f            | TABLE
(8 rows)
```

# SHOW MODEL

Muestra información útil sobre un modelo de machine learning, incluido el estado, los parámetros utilizados para crearlo y la función de predicción con los tipos de argumentos de entrada. Puede

utilizar la información de SHOW MODEL para volver a crear el modelo. Si se han modificado las tablas base, cuando se ejecuta CREATE MODEL con la misma instrucción SQL, se obtiene un modelo diferente. La información que devuelve SHOW MODEL es diferente para el propietario del modelo y para un usuario con el privilegio EXECUTE. SHOW MODEL muestra diferentes salidas cuando un modelo se forma en Amazon Redshift o cuando el modelo es un modelo BYOM.

## Sintaxis

```
SHOW MODEL ( ALL | model_name )
```

## Parámetros

ALL

Devuelve todos los modelos que el usuario puede utilizar y sus esquemas.

model_name

El nombre del modelo. El nombre del modelo en un esquema debe ser único.

## Notas de uso

El comando SHOW MODEL devuelve la siguiente información:

- El nombre del modelo.
- El esquema en el que se creó el modelo.
- El propietario del modelo.
- La hora de creación del modelo.
- El estado del modelo, como READY, TRAINING o FAILED.
- El mensaje de motivo de un modelo que presente error.
- El error de validación si el modelo finalizó la formación.
- El costo estimado necesario para derivar el modelo para un enfoque distinto de BYOM. Solo el propietario del modelo puede ver esta información.
- Una lista de parámetros especificados por el usuario y sus valores, concretamente los siguientes:
  - La columna TARGET especificada.
  - El tipo de modelo, AUTO o XGBoost.

- El tipo de problema, como REGRESSION, BINARY_CLASSIFICATION, MULTICLASS_CLASSIFICATION. Este parámetro es específico para AUTO.
- El nombre del trabajo de entrenamiento de Amazon SageMaker AI o del trabajo del Piloto automático de Amazon SageMaker AI que creó el modelo. Puede utilizar este nombre de trabajo para encontrar más información acerca del modelo en Amazon SageMaker AI.
- El objetivo, como MSE, F1, Accuracy. Este parámetro es específico para AUTO.
- El nombre de la función creada.
- El tipo de inferencia, local o remota.
- Los argumentos de entrada de la función de predicción.
- Los tipos de argumentos de entrada de la función de predicción para modelos que no empleen el enfoque Bring your own model (BYOM).
- El tipo de retorno de la función de predicción. Este parámetro es específico de BYOM.
- El nombre del punto de conexión de Amazon SageMaker AI para un modelo BYOM con inferencia remota.
- El rol de IAM. Solo el propietario del modelo puede ver esto.
- El bucket de S3 utilizado. Solo el propietario del modelo puede ver esto.
- La clave de AWS KMS, si se proporcionó una. Solo el propietario del modelo puede ver esto.
- El tiempo máximo que se puede ejecutar el modelo.
- Si el tipo de modelo no es AUTO, Amazon Redshift también muestra la lista de hiperparámetros proporcionados y sus valores.

También puede ver parte de la información proporcionada por SHOW MODEL en otras tablas de catálogo, como pg_proc. Amazon Redshift devuelve información sobre la función de predicción registrada en la tabla de catálogo pg_proc. Esta información incluye los nombres de los argumentos de entrada y los tipos para la función de predicción. Amazon Redshift devuelve la misma información en el comando SHOW MODEL.

```
SELECT * FROM pg_proc WHERE proname ILIKE '%<function_name>%';
```

## Ejemplos

En el siguiente ejemplo, se muestra la salida de SHOW MODEL.

```
SHOW MODEL ALL;
```

```
Schema Name |  Model Name
------------+---------------
 public     | customer_churn
```

El propietario de customer_churn puede ver la siguiente salida. Un usuario que tiene solo el privilegio EXECUTE no puede ver el rol de IAM, el bucket de Amazon S3 ni el costo estimado del modo.

```
SHOW MODEL customer_churn;

       Key                  |           Value
----------------------------+-----------------------------------
 Model Name                 | customer_churn
 Schema Name                | public
 Owner                      | 'owner'
 Creation Time              | Sat, 15.01.2000 14:45:20
 Model State                | READY
 validation:F1              | 0.855
 Estimated Cost             | 5.7
                            |
 TRAINING DATA:             |
 Table                      | customer_data
 Target Column              | CHURN
                            |
 PARAMETERS:                |
 Model Type                 | auto
 Problem Type               | binary_classification
 Objective                  | f1
 Function Name              | predict_churn
 Function Parameters        | age zip average_daily_spend average_daily_cases
 Function Parameter Types   | int int float float
 IAM Role                   | 'iam_role'
 KMS Key                    | 'kms_key'
 Max Runtime                | 36000
```

# SHOW DATASHARES

Muestra los datashares entrantes y salientes en un clúster desde la misma cuenta o entre cuentas. Si no especifica un nombre de datashare, Amazon Redshift mostrará todos los datashares de todas las bases de datos en el clúster. Los usuarios que tengan los privilegios ALTER y SHARE pueden ver los datashares para los que tienen privilegios.

### Sintaxis

```
SHOW DATASHARES [ LIKE 'namepattern' ]
```

### Parámetros

LIKE

Se trata de una cláusula opcional que compara el patrón de nombre especificado con la descripción del datashare. Cuando se utiliza esta cláusula, Amazon Redshift muestra solo los datashares con nombres que coinciden con el patrón de nombre especificado.

namepattern

Se trata del nombre del datashare solicitado o parte del nombre que coincidirá mediante caracteres comodín.

### Ejemplos

En el siguiente ejemplo, se muestran los datashares entrantes y salientes de un clúster.

```
SHOW DATASHARES;
SHOW DATASHARES LIKE 'sales%';

share_name   | share_owner | source_database | consumer_database | share_type |
 createdate          | is_publicaccessible | share_acl | producer_account |
 producer_namespace
-------------+-------------+-----------------+-------------------+------------
+---------------------+---------------------+-----------+------------------
+--------------------------------------
'salesshare' | 100         | dev             |                   | outbound
 | 2020-12-09 01:22:54.| False               |           |   123456789012   |
 13b8833d-17c6-4f16-8fe4-1a018f5ed00d
```

## SHOW PROCEDURE

Muestra la definición de un procedimiento almacenado dado, incluida su firma. Puede utilizar la salida de un SHOW PROCEDURE para recrear el procedimiento almacenado.

## Sintaxis

```
SHOW PROCEDURE sp_name [( [ [ argname ] [ argmode ] argtype [, ...] ] )]
```

## Parámetros

sp_name

El nombre del procedimiento que debe mostrarse.

[argname] [ argmode] argtype

Tipos de argumento de entrada para identificar el procedimiento almacenado. De manera opción, puede incluir los tipos de datos del argumento completo, incluidos los argumentos OUT. Esta parte es opcional si el nombre del procedimiento almacenado es único (esto es, no está sobrecargado).

## Ejemplos

En el siguiente ejemplo, se muestra la definición del procedimiento `test_spl2`.

```
show procedure test_sp2(int, varchar);
                                        Stored Procedure Definition
------------------------------------------------------------------------------------------------
CREATE OR REPLACE PROCEDURE public.test_sp2(f1 integer, INOUT f2 character varying, OUT
 character varying)
LANGUAGE plpgsql
AS $_$
DECLARE
out_var alias for $3;
loop_var int;
BEGIN
IF f1 is null OR f2 is null THEN
RAISE EXCEPTION 'input cannot be null';
END IF;
CREATE TEMP TABLE etl(a int, b varchar);
FOR loop_var IN 1..f1 LOOP
insert into etl values (loop_var, f2);
f2 := f2 || '+' || f2;
END LOOP;
SELECT INTO out_var count(*) from etl;
END;
```

```
$_$

(1 row)
```

# MOSTRAR ESQUEMAS

Muestra una lista de esquemas de una base de datos, junto con algunos atributos de esquema.

Cada fila de salida consta del nombre de base de datos, el nombre de esquema, el propietario de esquema, el tipo de esquema, la ACL de esquema, la base de datos de origen y la opción de esquema. Para obtener más información sobre estos atributos, consulte SVV_ALL_SCHEMAS.

Si el comando SHOW SCHEMAS puede dar como resultado más de 10 000 esquemas, se devuelve un error.

## Permisos necesarios

Para ver un esquema de una tabla de Amazon Redshift, el usuario actual debe cumplir uno de los siguientes criterios:

- Ser un superusuario.
- Ser el nuevo propietario del esquema.
- Tener concedido el privilegio USAGE en el esquema.

## Sintaxis

```
SHOW SCHEMAS FROM DATABASE database_name [LIKE 'filter_pattern'] [LIMIT row_limit ]
```

## Parámetros

database_name

Nombre de la base de datos que contiene las tablas que se enumerarán.

Para mostrar tablas en un AWS Glue Data Catalog, especifique (`awsdatacatalog`) como nombre de base de datos y asegúrese de que la configuración del sistema `data_catalog_auto_mount` está establecida a `true`. Para obtener más información, consulte ALTER SYSTEM.

filter_pattern

Expresión de caracteres UTF-8 válida con un patrón para hacer coincidir los nombres de esquema. La opción LIKE realiza una coincidencia que distingue entre mayúsculas y minúsculas y admite los siguientes metacaracteres de coincidencia de patrones:

| Metacarácter | Descripción |
|---|---|
| % | Coincide con cualquier secuencia de cero o más caracteres. |
| _ | Coincide con cualquier carácter. |

Si filter_pattern no contiene metacaracteres, solo representa la propia cadena; en ese caso, LIKE actúa igual que el operador de igualdad.

row_limit

Número máximo de filas que se devolverán. El valor de row_limit está entre 0 y 10 000.

## Ejemplos

En el siguiente ejemplo se muestran los esquemas de la base de datos de Amazon Redshift denominada `dev`.

```
SHOW SCHEMAS FROM DATABASE dev;

 database_name |     schema_name      | schema_owner | schema_type |         schema_acl
          | source_database | schema_option
---------------+----------------------+--------------+-------------
+----------------------------+-----------------+---------------
 dev           | pg_automv            |            1 | local       |
          |                 |
 dev           | pg_catalog           |            1 | local       | jpuser=UC/
jpuser~=U/jpuser  |                 |
 dev           | public               |            1 | local       | jpuser=UC/
jpuser~=UC/jpuser |                 |
 dev           | information_schema   |            1 | local       | jpuser=UC/
jpuser~=U/jpuser  |                 |
 dev           | schemad79cd6d93bf043 |            1 | local       |
          |                 |
```

En el siguiente ejemplo se muestran los esquemas de la base de datos de AWS Glue Data Catalog denominada `awsdatacatalog`. El número máximo de filas de salida es 5.

```
SHOW SCHEMAS FROM DATABASE awsdatacatalog LIMIT 5;

 database_name  |     schema_name      | schema_owner | schema_type | schema_acl |
 source_database | schema_option
----------------+----------------------+--------------+-------------+------------
+-----------------+---------------
 awsdatacatalog | 000_too_many_glue_db |              | EXTERNAL    |            |
            |
 awsdatacatalog | 123_default          |              | EXTERNAL    |            |
            |
 awsdatacatalog | adhoc                |              | EXTERNAL    |            |
            |
 awsdatacatalog | all_shapes_10mb      |              | EXTERNAL    |            |
            |
 awsdatacatalog | all_shapes_1g        |              | EXTERNAL    |            |
            |
```

# SHOW TABLE

Muestra la definición de una tabla, incluidos los atributos de tabla, las restricciones de tabla, los atributos de columna y las restricciones de columna. Puede utilizar la salida de la instrucción SHOW TABLE para recrear la tabla.

Para obtener más información acerca de la creación de tablas, consulte CREATE TABLE.

## Sintaxis

```
SHOW TABLE [schema_name.]table_name
```

## Parámetros

schema_name

(Opcional) Se trata del nombre del esquema relacionado.

table_name

Se trata del nombre de la tabla que se mostrará.

## Ejemplos

A continuación, se presenta un ejemplo de la salida de SHOW TABLE para la tabla `sales`.

```
show table sales;
```

```
CREATE TABLE public.sales (
salesid integer NOT NULL ENCODE az64,
listid integer NOT NULL ENCODE az64 distkey,
sellerid integer NOT NULL ENCODE az64,
buyerid integer NOT NULL ENCODE az64,
eventid integer NOT NULL ENCODE az64,
dateid smallint NOT NULL,
qtysold smallint NOT NULL ENCODE az64,
pricepaid numeric(8,2) ENCODE az64,
commission numeric(8,2) ENCODE az64,
saletime timestamp without time zone ENCODE az64
)
DISTSTYLE KEY SORTKEY ( dateid );
```

A continuación, se presenta un ejemplo de la salida de SHOW TABLE para la tabla `category` en el esquema `public`.

```
show table public.category;
```

```
CREATE TABLE public.category (
catid smallint NOT NULL distkey,
catgroup character varying(10) ENCODE lzo,
catname character varying(10) ENCODE lzo,
catdesc character varying(50) ENCODE lzo
) DISTSTYLE KEY SORTKEY ( catid );
```

En el siguiente ejemplo, se crea la tabla `foo` con una clave principal.

```
create table foo(a int PRIMARY KEY, b int);
```

Los resultados de SHOW TABLE muestran la instrucción de creación con todas las propiedades de la tabla `foo`.

```
show table foo;
```

```
CREATE TABLE public.foo ( a integer NOT NULL ENCODE az64, b integer ENCODE az64,
 PRIMARY KEY (a) ) DISTSTYLE AUTO;
```

# SHOW TABLES

Muestra una lista de tablas de un esquema, junto con algunos atributos de tabla.

Cada fila de salida consta del nombre de base de datos, el nombre de esquema, el nombre de tabla, el tipo de tabla, la ACL de tabla y las observaciones. Para obtener más información sobre estos atributos, consulte SVV_ALL_TABLES.

Si el comando SHOW TABLES diera como resultado más de 10 000 tablas, se devolvería un error.

## Permisos necesarios

Para ver una tabla de un esquema de Amazon Redshift, el usuario actual debe cumplir uno de los siguientes criterios:

- Ser un superusuario.
- Ser el propietario de la tabla.
- Tener concedido el privilegio USAGE en el esquema principal y el privilegio SELECT en la tabla o el privilegio SELECT en cualquier columna de la tabla.

## Sintaxis

```
SHOW TABLES FROM SCHEMA database_name.schema_name [LIKE 'filter_pattern']
 [LIMIT row_limit ]
```

## Parámetros

database_name

Nombre de la base de datos que contiene las tablas que se enumerarán.

Para mostrar tablas en un AWS Glue Data Catalog, especifique (`awsdatacatalog`) como nombre de base de datos y asegúrese de que la configuración del sistema `data_catalog_auto_mount` está establecida a `true`. Para obtener más información, consulte ALTER SYSTEM.

schema_name

Nombre del esquema que contiene las tablas que se enumerarán.

Para mostrar las tablas de AWS Glue Data Catalog, indique el nombre de la base de datos de AWS Glue como nombre del esquema.

filter_pattern

Expresión de caracteres UTF-8 válida con un patrón para hacer coincidir los nombres de tabla. La opción LIKE realiza una coincidencia que distingue entre mayúsculas y minúsculas y admite los siguientes metacaracteres de coincidencia de patrones:

| Metacarácter | Descripción |
|---|---|
| % | Coincide con cualquier secuencia de cero o más caracteres. |
| _ | Coincide con cualquier carácter. |

Si filter_pattern no contiene metacaracteres, solo representa la propia cadena; en ese caso, LIKE actúa igual que el operador de igualdad.

row_limit

Número máximo de filas que se devolverán. El valor de row_limit está entre 0 y 10 000.

## Ejemplos

En el siguiente ejemplo se muestran las tablas de la base de datos de Amazon Redshift denominada `dev` que se encuentran en el esquema `public`.

```
SHOW TABLES FROM SCHEMA dev.public;

 database_name | schema_name | table_name | table_type | table_acl | remarks
---------------+-------------+------------+------------+-----------+---------
 dev           | public      | tb         | TABLE      |           |
 dev           | public      | tb2        | TABLE      |           |
 dev           | public      | tb3        | TABLE      |           |
```

En el siguiente ejemplo se muestran las columnas de la base de datos de AWS Glue Data Catalog denominada `awsdatacatalog` que se encuentran en el esquema `batman`.

```
SHOW TABLES FROM SCHEMA awsdatacatalog.batman;

 database_name  | schema_name |    table_name    | table_type | table_acl | remarks
----------------+-------------+------------------+------------+-----------+---------
 awsdatacatalog | batman      | nation           | EXTERNAL   |           |
 awsdatacatalog | batman      | part             | EXTERNAL   |           |
 awsdatacatalog | batman      | partsupp         | EXTERNAL   |           |
 awsdatacatalog | batman      | region           | EXTERNAL   |           |
 awsdatacatalog | batman      | supplier         | EXTERNAL   |           |
 awsdatacatalog | batman      | automount_nation | EXTERNAL   |           |
```

# SHOW VIEW

Muestra la definición de una vista, incluidas las vistas materializadas y las vistas de enlace de tiempo de ejecución. Puede utilizar la salida de la instrucción SHOW VIEW para recrear la vista.

## Sintaxis

```
SHOW VIEW [schema_name.]view_name
```

## Parámetros

schema_name

(Opcional) Se trata del nombre del esquema relacionado.

view_name

Se trata del nombre de la vista que se mostrará.

## Ejemplos

A continuación, se presenta la definición de la vista `LA_Venues_v`.

```
create view LA_Venues_v as select * from venue where venuecity='Los Angeles';
```

A continuación, se presenta un ejemplo del comando SHOW VIEW y la salida para la vista definida anteriormente.

```
show view LA_Venues_v;
```

```
SELECT venue.venueid,
venue.venuename,
venue.venuecity,
venue.venuestate,
venue.venueseats
FROM venue WHERE ((venue.venuecity)::text = 'Los Angeles'::text);
```

A continuación, se presenta la definición de la vista `public.Sports_v` en el esquema `public`.

```
create view public.Sports_v as select * from category where catgroup='Sports';
```

A continuación, se presenta un ejemplo del comando SHOW VIEW y la salida para la vista definida anteriormente.

```
show view public.Sports_v;
```

```
SELECT category.catid,
category.catgroup,
category.catname,
category.catdesc
FROM category WHERE ((category.catgroup)::text = 'Sports'::text);
```

## START TRANSACTION

Sinónimo de la función BEGIN.

Consulte BEGIN.

## TRUNCATE

Elimina todas las filas de una tabla sin realizar un análisis de la tabla: esta operación es una alternativa más rápida a una operación DELETE no calificada. Para ejecutar un comando TRUNCATE, debe tener el permiso TRUNCATE de la tabla, ser el propietario de la tabla o ser un superusuario. Para conceder permisos para truncar una tabla, utilice el comando GRANT.

TRUNCATE es mucho más eficiente que DELETE y no requiere VACUUM ni ANALYZE. No obstante, tenga en cuenta que TRUNCATE confirma la transacción en la que se ejecuta.

## Sintaxis

```
TRUNCATE [ TABLE ] table_name
```

El comando también funciona en una vista materializada.

```
TRUNCATE materialized_view_name
```

## Parámetros

TABLE

Palabra clave opcional.

table_name

Una tabla temporal o persistente. Solo el propietario de la tabla o un superusuario puede truncarla.

Puede truncar cualquier tabla, incluidas las tablas a las que se hace referencia en limitaciones de clave externa.

No es necesario limpiar una tabla después de truncarla.

materialized_view_name

Una vista materializada.

Puede truncar una vista materializada que se utilice para Ingesta de streaming a una vista materializada.

## Notas de uso

El comando TRUNCATE confirma la transacción en la que se ejecuta; por lo tanto, no se puede revertir una operación TRUNCATE, y un comando TRUNCATE puede confirmar otras operaciones cuando se confirma a sí mismo.

## Ejemplos

Utilice el comando TRUNCATE para eliminar todas las filas de la tabla CATEGORY:

```
truncate category;
```

Intento de revertir una operación TRUNCATE:

```
begin;

truncate date;

rollback;

select count(*) from date;
count
-------
0
(1 row)
```

La tabla DATE permanece vacía después del comando ROLLBACK ya que el comando TRUNCATE se confirmó automáticamente.

En el siguiente ejemplo, se utiliza el comando TRUNCATE para eliminar todas las filas de una vista materializada.

```
truncate my_materialized_view;
```

Elimina todos los registros de la vista materializada y deja intactos la vista materializada y su esquema. En la consulta, el nombre de la vista materializada es un ejemplo.

## UNLOAD

El cifrado del cliente para los comandos COPY y UNLOAD dejará de estar abierto a nuevos clientes a partir del 30 de abril de 2025. Si ha utilizado el cifrado del cliente con los comandos COPY y UNLOAD en los 12 meses anteriores al 30 de abril de 2025, puede seguir utilizando el cifrado del cliente con los comandos COPY o UNLOAD hasta el 30 de abril de 2026. Después del 30 de abril de 2026, no podrá utilizar el cifrado del cliente para COPY y UNLOAD. Le recomenda mos que pase a utilizar el cifrado del servidor para COPY y UNLOAD lo antes posible. Si ya está utilizando el cifrado del servidor para COPY y UNLOAD, no hay ningún cambio y puede seguir utilizándolo sin alterar las consultas. Para obtener más información sobre el cifrado para COPY y UNLOAD, consulte el parámetro ENCRYPTED a continuación.

Descarga el resultado de una consulta en uno o más archivos de texto, JSON o Apache Parquet en Amazon S3 con cifrado del lado del servidor (SSE-S3) de Amazon S3. También puede especificar el cifrado del servidor con una clave de AWS Key Management Service (SSE-KMS).

De manera predeterminada, el formato del archivo descargado es texto delimitado por barras verticales ( | ).

Puede administrar el tamaño de los archivos en Amazon S3 y, por extensión, la cantidad de archivos mediante la configuración del parámetro MAXFILESIZE. Asegúrese de que los rangos de IP de S3 estén agregados a la lista de permitidos. Para obtener más información acerca de los rangos de IP de S3 necesarios, consulte Aislamiento de red.

Puede descargar el resultado de una consulta de Amazon Redshift en su lago de datos de Amazon S3 en Apache Parquet, un formato de almacenamiento en columnas, abierto y eficiente para análisis. El formato Parquet es hasta 2 veces más rápido de descargar y consume hasta 6 veces menos almacenamiento en Amazon S3, en comparación con los formatos de texto. Esto le permite guardar la transformación y el enriquecimiento de datos que haya realizado en Amazon S3, en el lago de datos de Amazon S3 en un formato abierto. Luego, puede analizar los datos con Redshift Spectrum y otros servicios de AWS, como Amazon Athena, Amazon EMR y Amazon SageMaker AI.

Para obtener más información y ejemplos de escenarios sobre el uso del comando UNLOAD, consulte Descarga de datos en Amazon Redshift.

## Privilegios y permisos necesarios

Para que el comando UNLOAD se ejecute correctamente, se necesita al menos el privilegio SELECT en los datos de la base de datos, junto con el permiso para escribir en la ubicación de Amazon S3. Para obtener información sobre los permisos de acceso a los recursos de AWS para el comando UNLOAD, consulte Permisos para acceder a otros recursos de AWS.

Para aplicar permisos con privilegios mínimos, siga estas recomendaciones para conceder únicamente los permisos necesarios al usuario que ejecute el comando.

- El usuario también debe tener el privilegio SELECT sobre los datos. Para obtener información acerca de cómo limitar los privilegios en las bases de datos, consulte GRANT.
- El usuario necesita permiso para asumir el rol de IAM para escribir en el bucket de Amazon S3 de su Cuenta de AWS. Para restringir el acceso a un usuario de la base de datos para que asuma un rol, consulte Restricción de acceso a los roles de IAM en la Guía de administración de Amazon Redshift.

- El usuario necesita tener acceso al bucket de Amazon S3. Para restringir los permisos mediante una política de bucket de Amazon S3, consulte Políticas de bucket para Amazon S3 en la Guía del usuario de Amazon Simple Storage Service.

## Sintaxis

```
UNLOAD ('select-statement')
TO 's3://object-path/name-prefix'
authorization
[ option, ...]

where authorization is
IAM_ROLE { default | 'arn:aws:iam::<Cuenta de AWS-id-1>:role/<role-
name>[,arn:aws:iam::<Cuenta de AWS-id-2>:role/<role-name>][,...]' }

where option is
| [ FORMAT [ AS ] ] CSV | PARQUET | JSON
| PARTITION BY ( column_name [, ... ] ) [ INCLUDE ]
| MANIFEST [ VERBOSE ]
| HEADER
| DELIMITER [ AS ] 'delimiter-char'
| FIXEDWIDTH [ AS ] 'fixedwidth-spec'
| ENCRYPTED [ AUTO ]
| BZIP2
| GZIP
| ZSTD
| ADDQUOTES
| NULL [ AS ] 'null-string'
| ESCAPE
| ALLOWOVERWRITE
| CLEANPATH
| PARALLEL [ { ON | TRUE } | { OFF | FALSE } ]
| MAXFILESIZE [AS] max-size [ MB | GB ]
| ROWGROUPSIZE [AS] size [ MB | GB ]
| REGION [AS] 'aws-region' }
| EXTENSION 'extension-name'
```

## Parámetros

('select-statement [instrucción-select]')

Una consulta SELECT. Se descargan los resultados de la consulta. En la mayoría de casos, es útil descargar datos en un orden al especificar una cláusula ORDER BY en la consulta. Este enfoque permite ahorrar el tiempo que se requiere para ordenar los datos cuando se vuelven a cargar.

La consulta debe escribirse entre comillas simples, como se muestra a continuación:

```
('select * from venue order by venueid')
```

> **Note**
>
> Si la consulta contiene comillas (por ejemplo, para encerrar valores literales), coloque el literal entre dos conjuntos de comillas simples. También debe incluir la consulta entre comillas simples:
>
> ```
> ('select * from venue where venuestate=''NV''')
> ```

TO 's3://object-path/name-prefix'

Se trata de la ruta completa, incluido el nombre del bucket, a la ubicación en Amazon S3 donde Amazon Redshift escribe los objetos del archivo de salida, incluido el archivo de manifiesto si se especifica MANIFEST. Los nombres del objeto tienen el prefijo name-prefix. Si utiliza `PARTITION BY`, se agrega automáticamente una barra diagonal (/) al final del valor nombre-prefijo, si es necesario. Para mayor seguridad, UNLOAD se conecta a Amazon S3 con una conexión HTTPS. Por defecto, UNLOAD escribe uno o más archivos por sector. UNLOAD agrega un número de sector y un número de parte al prefijo del nombre especificado, como se muestra a continuación:

*<object-path>*/*<name-prefix>*<*slice-number*>_part_*<part-number>*.

Si se especifica MANIFEST, el archivo de manifiesto se escribe de la siguiente manera:

*<object_path>*/*<name_prefix>*manifest.

Si se especifica PARALLEL en OFF, los archivos de datos se escriben de la siguiente manera:

*<object_path>*/*<name_prefix>*<*part-number*>.

UNLOAD crea de forma automática archivos cifrados a través del cifrado del lado del servidor (SSE) de Amazon S3, incluido el archivo de manifiesto si se utiliza MANIFEST. El comando COPY lee automáticamente los archivos cifrados del servidor durante la operación de carga. Puede descargar de manera transparente los archivos cifrados del lado del servidor desde el bucket a través de la API o la consola de Amazon S3. Para obtener más información, consulte Protección de los datos con el cifrado del servidor.

> **Important**
>
> REGION es necesario cuando el bucket de Amazon S3 no está en la misma región de Región de AWS que la base de datos de Amazon Redshift.

authorization

El comando UNLOAD necesita autorización para escribir datos en Amazon S3. El comando UNLOAD usa los mismos parámetros que el comando COPY usa para la autorización. Para obtener más información, consulte Parámetros de autorización en la referencia de sintaxis del comando COPY.

IAM_ROLE { default | 'arn:aws:iam::*<Cuenta de AWS-id-1>*:role/*<role-name>*'

Utilice la palabra clave predeterminada para que Amazon Redshift utilice el rol de IAM configurado como predeterminado y asociado al clúster cuando se ejecuta el comando UNLOAD.

Utilice el nombre de recurso de Amazon (ARN), de un rol de IAM que el clúster utiliza para la autenticación y la autorización. Si especifica IAM_ROLE, no puede utilizar ACCESS_KEY_ID, SECRET_ACCESS_KEY, SESSION_TOKEN ni CREDENTIALS. El IAM_ROLE se puede encadenar. Para obtener más información, consulte Encadenar roles de IAM en la Guía de administración de Amazon Redshift.

[ FORMAT [AS] ] CSV | PARQUET | JSON

Se trata de las palabras clave para especificar el formato de descarga con objeto de anular el formato predeterminado.

Cuando CSV, descarga en un archivo de texto en formato CSV utilizando un carácter de coma ( , ) como delimitador predeterminado. Si un campo contiene delimitadores, comillas dobles,

caracteres de nueva línea o retornos de carro, el campo en el archivo descargado se encuentra entre comillas dobles. Se aplica escape a un carácter de comillas dobles dentro de un campo de datos mediante un carácter de comillas dobles adicional. Cuando se descargan cero filas, es posible que Amazon Redshift escriba objetos de Amazon S3 vacíos.

Si es PARQUET, se descarga en un archivo en formato de Apache Parquet versión 1.0. De manera predeterminada, cada grupo de filas se comprime mediante la compresión SNAPPY. Para obtener más información sobre el formato de Apache Parquet, consulte Parquet.

Si es JSON, se descarga en un archivo JSON con cada línea que contiene un objeto JSON, lo que representa un registro completo en el resultado de la consulta. Amazon Redshift admite escribir JSON anidado cuando el resultado de la consulta contiene columnas SUPER. Para crear un objeto JSON válido, el nombre de cada columna de la consulta debe ser único. En el archivo JSON, los valores booleanos se descargan como `t` o `f`, y los valores NULL se descargan como `null`. Cuando se descargan cero filas, Amazon Redshift no escribe objetos de Amazon S3.

Las palabras clave FORMAT y AS son opcionales. No puede usar CSV con ESCAPE, FIXEDWIDTH ni ADDQUOTES. No puede utilizar PARQUET con DELIMITER, FIXEDWIDTH, ADDQUOTES, ESCAPE, NULL AS, HEADER, GZIP, BZIP2 o ZSTD. PARQUET con ENCRYPTED solo se admite con el cifrado en el servidor con una clave AWS Key Management Service (SSE-KMS). No puede utilizar JSON con DELIMITER, HEADER, FIXEDWIDTH, ADDQUOTES, ESCAPE, o NULL AS.

PARTITION BY ( column_name [, ... ] ) [INCLUDE]

Especifica las claves de partición para la operación de descarga. UNLOAD automáticamente particiona los archivos de salida en carpetas de partición basándose en los valores de las claves de partición, siguiendo la convención de la Apache Hive. Por ejemplo, un archivo Parquet que pertenece al año de partición 2019 y al mes de septiembre tiene el siguiente prefijo: `s3://amzn-s3-demo-bucket/my_prefix/year=2019/month=September/000.parquet`.

El valor de column_name debe ser una columna en los resultados de consulta que se están descargando.

Si especifica PARTITION BY con la opción INCLUDE, las columnas de partición no se quitarán de los archivos descargados.

Amazon Redshift no admite literales de cadena en cláusulas PARTITION BY.

MANIFEST [ VERBOSE ]

Crea un archivo de manifiesto que enumera explícitamente detalles de los archivos de datos que son creados por el proceso UNLOAD. El manifiesto es un archivo de texto en formato JSON que enumera las direcciones URL de cada archivo escrito en Amazon S3.

Si se especifica MANIFEST con la opción VERBOSE, el manifiesto incluye los siguientes detalles:

- Los nombres de columnas y los tipos de datos y, para los tipos de datos CHAR, VARCHAR o NUMERIC, las dimensiones para cada columna. Para los tipos de datos CHAR y VARCHAR, la dimensión es la longitud. Para un tipo de datos DECIMAL o NUMERIC, las dimensiones son la precisión y la escala.
- La cantidad de filas descargadas en cada archivo. Si se especifica la opción HEADER, la cantidad de filas incluye la línea de encabezado.
- El tamaño de archivo total de todos los archivos descargados y el recuento de filas total descargadas en todos los archivos. Si se especifica la opción HEADER, el recuento de filas incluye las líneas de encabezado.
- El autor. El autor es siempre "Amazon Redshift".

Solo puede especificar VERBOSE después de MANIFEST.

El archivo de manifiesto se escribe en el mismo prefijo de ruta de Amazon S3 que los archivos de descarga en el formato `<object_path_prefix>manifest`. Por ejemplo, si UNLOAD especifica el prefijo de ruta “`s3://amzn-s3-demo-bucket/venue_`” de Amazon S3, la ubicación del archivo de manifiesto es “`s3://amzn-s3-demo-bucket/venue_manifest`”.

HEADER

Agrega una línea de encabezado que contiene los nombres de columna situados al comienzo de cada archivo de salida. Las opciones de transformación de texto, como CSV, DELIMITER, ADDQUOTES y ESCAPE, también se aplican a la línea de encabezado. No puede usar HEADER con FIXEDWIDTH.

DELIMITER AS 'delimiter_character'

Especifica un carácter ASCII único que se utiliza para separar campos en el archivo de salida, como un carácter de barra vertical ( | ), una coma ( , ) o una tabulación ( \t ). El delimitador predeterminado para los archivos de texto es un carácter de barra vertical. El delimitador predeterminado para los archivos CSV es un carácter de coma. La palabra clave AS es opcional. No puede usar DELIMITER con FIXEDWIDTH. Si los datos contienen el carácter delimitador,

necesita especificar la opción ESCAPE para aplicar escape al delimitador, o utilizar ADDQUOTES para encerrar los datos entre comillas dobles. También puede especificar un delimitador que no esté contenido en los datos.

FIXEDWIDTH 'fixedwidth_spec'

Descarga los datos a un archivo donde el ancho de cada columna tiene una longitud fija, en lugar de estar separadas por un delimitador. La cadena fixedwidth_spec especifica la cantidad de columnas y el ancho de las mismas. La palabra clave AS es opcional. Como FIXEDWIDTH no trunca los datos, la especificación de cada columna en la instrucción UNLOAD debe tener al menos la misma longitud que la entrada más larga de esa columna, A continuación, se muestra el formato de fixedwidth_spec:

```
'colID1:colWidth1,colID2:colWidth2, ...'
```

No puede usar FIXEDWIDTH con DELIMITER o HEADER.

ENCRYPTED [AUTO]

Especifica que los archivos de salida en Amazon S3 se cifran mediante el cifrado del servidor de Amazon S3. Si se especifica MANIFEST, también se cifra el archivo de manifiesto. Para obtener más información, consulte Descarga de archivos de datos cifrados. Si no especifica el parámetro ENCRYPTED, UNLOAD crea de forma automática archivos cifrados a través del cifrado del lado del servidor de Amazon S3, con claves de cifrado administradas por AWS (SSE-S3).

Para ENCRYPTED, es posible que desee descargar el contenido en Amazon S3 mediante el cifrado del lado del servidor con una clave de AWS KMS (SSE-KMS). Si es así, utilice el parámetro KMS_KEY_ID para proporcionar el ID de la clave. No puede utilizar el parámetro the section called “Uso de CREDENTIALS” con el parámetro KMS_KEY_ID. Si ejecuta un comando UNLOAD para datos mediante KMS_KEY_ID, puede realizar una operación COPY para los mismos datos sin especificar una clave.

Si se utiliza ENCRYPTED AUTO, el comando UNLOAD obtiene la clave de cifrado de AWS KMS predeterminada en la propiedad del bucket de Amazon S3 de destino y cifra los archivos escritos en Amazon S3 con la clave de AWS KMS. Si el bucket no tiene la clave de cifrado de AWS KMS predeterminada, UNLOAD crea automáticamente archivos cifrados a través del cifrado del lado del servidor de Amazon Redshift con claves de cifrado administradas por AWS (SSE-S3). No puede utilizar esta opción con KMS_KEY_ID, MASTER_SYMMETRIC_KEY ni CREDENTIALS que contiene master_symmetric_key.

KMS_KEY_ID 'key-id'

Especifica el ID de una clave de AWS Key Management Service (AWS KMS) que se utilizará para cifrar archivos de datos en Amazon S3. Para obtener más información, consulte ¿Qué es AWS Key Management Service? Si especifica KMS_KEY_ID, también debe especificar el parámetro ENCRYPTED. Si especifica KMS_KEY_ID, no puede autenticar con el parámetro CREDENTIALS. En su lugar, utilice the section called “Uso de IAM_ROLE” o the section called “Uso de ACCESS_KEY_ID y SECRET_ACCESS_KEY”.

BZIP2

Descarga datos en uno o más archivos comprimidos en bzip2 por sector. Cada archivo resultante se adjunta con una extensión `.bz2`.

GZIP

Descarga datos en uno o más archivos comprimidos en gzip por sector. Cada archivo resultante se adjunta con una extensión `.gz`.

ZSTD

Descarga datos en uno o más archivos comprimidos en Zstandard por sector. Cada archivo resultante se adjunta con una extensión `.zst`.

ADDQUOTES

Coloca comillas alrededor de cada campo de datos descargado para que Amazon Redshift pueda descargar los valores de datos que contienen el propio delimitador. Por ejemplo, si el delimitador es una coma, podría descargar y volver a cargar los siguientes datos de manera correcta:

```
"1","Hello, World"
```

Sin las comillas agregadas, la cadena `Hello, World` se analizaría como dos campos separados.

Algunos formatos de salida no admiten ADDQUOTES.

Si utiliza ADDQUOTES, debe especificar REMOVEQUOTES en el comando COPY si vuelve a cargar los datos.

NULL AS 'null-string'

Especifica una cadena que representa un valor nulo en archivos de descarga. Si se utiliza esta opción, todos los archivos de salida contienen la cadena especificada en lugar de valores nulos

encontrados en los datos seleccionados. Si no se especifica esta opción, los valores nulos se descargan como:

- Cadenas de longitud cero para un resultado delimitado
- Cadenas con espacios en blanco para un resultado de ancho fijo

Si se especifica una cadena nula para una descarga de ancho fijo, y el ancho de una columna de resultado es menor que el ancho de la cadena nula, ocurre el siguiente comportamiento:

- Un campo vacío es el resultado para columnas sin caracteres
- Se reporta un error para columnas con caracteres

A diferencia de otros tipos de datos en los que una cadena definida por el usuario representa un valor nulo, Amazon Redshift exporta las columnas de datos SUPER mediante el formato JSON y lo representa como nulo, de acuerdo con lo determinado por el formato JSON. Como resultado, las columnas de datos SUPER ignoran la opción NULL [AS] utilizada en los comandos UNLOAD.

ESCAPE

Para las columnas CHAR y VARCHAR en archivos de descarga delimitados, se coloca un carácter de escape (\) antes de cada ocurrencia de los siguientes caracteres:

- Salto de línea: `\n`
- Retorno de carro: `\r`
- El carácter delimitador especificado para los datos descargados.
- El carácter de escape: `\`
- Un carácter de comillas: `"` o `'` (si se especifican ESCAPE y ADDQUOTES en el comando UNLOAD).

> **Important**
>
> Si cargó los datos a través de un comando COPY con la opción ESCAPE, también debe especificar la opción ESCAPE con el comando UNLOAD para generar el archivo de salida recíproco. De manera similar, si utiliza UNLOAD con la opción ESCAPE, necesita usar ESCAPE cuando aplica COPY en los mismos datos.

ALLOWOVERWRITE

Por defecto, UNLOAD produce un error si encuentra archivos que podría sobrescribir. Si se especifica ALLOWOVERWRITE, UNLOAD sobrescribe los archivos existentes, incluido el archivo de manifiesto.

CLEANPATH

La opción CLEANPATH quita los archivos existentes ubicados en la ruta de Amazon S3 especificada en la cláusula TO antes de descargar archivos en la ubicación determinada.

Si incluye la cláusula PARTITION BY, los archivos existentes se quitan solo de las carpetas de partición para recibir archivos nuevos generados por la operación UNLOAD.

Debe tener el permiso `s3:DeleteObject` en el bucket de Amazon S3. Para obtener información, consulte Políticas y permisos en Amazon S3 en la Guía del usuario de Amazon Simple Storage Service. Los archivos que quita con la opción CLEANPATH se eliminan de forma permanente y no se pueden recuperar. Si el bucket de Amazon S3 de destino tiene activado el control de versiones, UNLOAD con la opción CLEANPATH no elimina las versiones anteriores de los archivos.

No se puede especificar la opción CLEANPATH si ya se especificó la opción ALLOWOVERWRITE.

PARALLEL

Por defecto, el comando UNLOAD escribe datos simultáneamente en varios archivos, según la cantidad de sectores que haya en el clúster. La opción predeterminada es ON o TRUE. Si PARALLEL es OFF o FALSE, UNLOAD escribe en uno o más archivos de datos en serie, ordenados absolutamente en función de la cláusula ORDER BY, si se utiliza una. El tamaño máximo de un archivo de datos es de 6,2 GB. Por ejemplo, si descarga 13,4 GB de datos, UNLOAD crea los siguientes tres archivos.

```
s3://amzn-s3-demo-bucket/key000    6.2 GB
s3://amzn-s3-demo-bucket/key001    6.2 GB
s3://amzn-s3-demo-bucket/key002    1.0 GB
```

> **Note**
>
> El comando UNLOAD está diseñado para utilizar procesamiento en paralelo. Le recomendamos dejar PARALLEL habilitado para la mayoría de casos, especialmente si los archivos se utilizan para cargar tablas mediante un comando COPY.

MAXFILESIZE [AS] max-size [ MB | GB ]

Especifica el tamaño máximo de los archivos que UNLOAD crea en Amazon S3. Especifique un valor decimal entre 5 MB y 6,2 GB. La palabra clave AS es opcional. La unidad predeterminada es MB. Si no se especifica MAXFILESIZE, el tamaño máximo predeterminado de un archivo es de 6,2 GB. El tamaño del archivo de manifiesto, si se utiliza uno, no se ve afectado por MAXFILESIZE.

ROWGROUPSIZE [AS] size [ MB | GB ]

Especifica el tamaño de los grupos de filas. La elección de un tamaño mayor puede reducir el número de grupos de filas y reducir la cantidad de comunicación de red. Especifique un valor entero entre 32 MB y 128 MB. La palabra clave AS es opcional. La unidad predeterminada es MB.

Si no se especifica ROWGROUPSIZE, el tamaño predeterminado es 32 MB. Para utilizar este parámetro, el formato de almacenamiento debe ser Parquet y el tipo de nodo debe ser ra3.4xlarge, ra3.16xlarge o dc2.8xlarge.

REGION [AS] 'aws_region'

Especifica la Región de AWS donde está ubicado el bucket de Amazon S3 de destino. REGION es obligatorio para la operación UNLOAD en un bucket de Amazon S3 que no esté en la misma Región de AWS que la base de datos de Amazon Redshift.

El valor de aws_region debe coincidir con una región de AWS incluida en la tabla Regiones y puntos de conexión de Amazon Redshift en la Referencia general de AWS.

De manera predeterminada, UNLOAD asume que el bucket de Amazon S3 de destino se encuentra en la misma Región de AWS que la base de datos de Amazon Redshift.

EXTENSION “extension-name”

Especifica la extensión del archivo que se va a agregar a los nombres de los archivos descargados. Amazon Redshift no ejecuta ninguna validación, por lo que debe comprobar que la extensión de archivo especificada es correcta. Si especifica un método de compresión sin

proporcionar una extensión, Amazon Redshift solo agrega la extensión del método de compresión al nombre del archivo. Si no proporciona ninguna extensión y no especifica un método de compresión, Amazon Redshift no añade nada al nombre del archivo.

## Notas de uso

### Uso de ESCAPE para todas las operaciones UNLOAD de texto delimitado

Cuando utiliza UNLOAD con un delimitador, sus datos pueden incluir ese delimitador o cualquiera de los caracteres enumerados en la descripción de la opción ESCAPE. En este caso, debe usar la opción ESCAPE con la instrucción UNLOAD. Si no utiliza la opción ESCAPE con UNLOAD, las operaciones de COPY subsiguientes que utilicen datos descargados producirán un error.

> **Important**
>
> Le recomendamos encarecidamente que siempre use ESCAPE con ambas instrucciones UNLOAD y COPY. Solo puede saltarse esta regla si sabe con seguridad que los datos no contienen delimitadores u otros caracteres que puedan necesitar escape.

### Pérdida de precisión de punto flotante

Puede enfrentarse a una pérdida de precisión para datos de punto flotante que se descargan y vuelven a cargar de manera sucesiva.

### Cláusula de límite

La consulta SELECT no puede utilizar una cláusula LIMIT en SELECT externo. Por ejemplo, la siguiente instrucción UNLOAD produce un error.

```
unload ('select * from venue limit 10')
to 's3://amzn-s3-demo-bucket/venue_pipe_' iam_role 'arn:aws:iam::0123456789012:role/
MyRedshiftRole';
```

En su lugar, use una cláusula LIMIT anidada, como en el siguiente ejemplo.

```
unload ('select * from venue where venueid in
(select venueid from venue order by venueid desc limit 10)')
to 's3://amzn-s3-demo-bucket/venue_pipe_' iam_role 'arn:aws:iam::0123456789012:role/
MyRedshiftRole';
```

También puede rellenar una tabla con SELECT…INTO o CREATE TABLE AS a través de una cláusula LIMIT y, luego, descargar desde esa tabla.

Descarga de una columna de tipo de datos GEOMETRY

Solo puede descargar columnas GEOMETRY para texto o formato CSV. No puede descargas datos GEOMETRY con la opción `FIXEDWIDTH`. Los datos se descargan en forma hexadecimal del formato binario conocido extendido (EWKB). Si el tamaño de los datos EWKB es mayor de 4 MB, se producirá una advertencia porque los datos no se podrán luego cargar en una tabla.

Descarga del tipo de datos HLLSKETCH

Solo puede descargar columnas HLLSKETCH en formato de texto o CSV. No puede descargar datos HLLSKETCH con la opción `FIXEDWIDTH`. Los datos se descargan en el formato Base64 para bocetos densos de HyperLogLog o en el formato JSON para bocetos dispersos de HyperLogLog. Para obtener más información, consulte Funciones HyperLogLog.

En el siguiente ejemplo, se exporta una tabla que contiene columnas HLLSKETCH a un archivo.

```
CREATE TABLE a_table(an_int INT, b_int INT);
INSERT INTO a_table VALUES (1,1), (2,1), (3,1), (4,1), (1,2), (2,2), (3,2), (4,2),
 (5,2), (6,2);

CREATE TABLE hll_table (sketch HLLSKETCH);
INSERT INTO hll_table select hll_create_sketch(an_int) from a_table group by b_int;

UNLOAD ('select * from hll_table') TO 's3://amzn-s3-demo-bucket/unload/'
IAM_ROLE 'arn:aws:iam::0123456789012:role/MyRedshiftRole' NULL AS 'null' ALLOWOVERWRITE
 CSV;
```

Descarga de una columna de tipo de datos VARBYTE

Solo puede descargar columnas VARBYTE para formato de texto o CSV. Los datos se descargan en forma hexadecimal. No se pueden descargar datos VARBYTE con la opción `FIXEDWIDTH`. No se admite la opción `ADDQUOTES` para DESCARGAR en un CSV. Una columna VARBYTE no se puede utilizar como columna PARTITIONED BY.

Cláusula FORMAT AS PARQUET

Tenga en cuenta estas consideraciones al usar FORMAT AS PARQUET:

- La descarga en Parquet no utiliza la compresión de nivel de archivo. Cada grupo de filas se comprime con SNAPPY.
- Si no se especifica MAXFILESIZE, el tamaño máximo predeterminado de un archivo es de 6,2 GB. Puede utilizar MAXFILESIZE para especificar el tamaño de un archivo de entre 5 MB y 6,2 GB. El tamaño real del archivo es aproximado cuando se está escribiendo el archivo, por lo que es posible que no sea exactamente igual al número que especifique.

  Para maximizar el rendimiento del análisis, Amazon Redshift intenta crear archivos Parquet que contengan grupos de filas de 32 MB del mismo tamaño. El valor MAXFILESIZE que especifique se redondea automáticamente al múltiplo más cercano de 32 MB. Por ejemplo, si especifica MAXFILESIZE 200 MB, cada archivo Parquet descargado es de aproximadamente 192 MB (grupo de filas de 32 MB x 6 = 192 MB).
- Si una columna utiliza el formato de datos TIMESTAMPTZ, solo se descargan los valores de marca temporal. La información de zona horaria no se descarga.
- No especifique prefijos de nombre de archivo que comience por los caracteres de subrayado (_) o punto (.). Redshift Spectrum trata los archivos que comienzan por estos caracteres como archivos ocultos y los ignora.

### Cláusula PARTITION BY

Tenga en cuenta estas consideraciones al usar PARTITION BY:

- Las columnas de partición no se incluyen en el archivo de salida.
- Asegúrese de incluir las columnas de partición en la consulta SELECT utilizada en la instrucción UNLOAD. Puede especificar cualquier número de columnas de partición en el comando UNLOAD. No obstante, existe una limitación de que debe haber al menos una columna que no sea de partición para formar parte del archivo.
- Si el valor de clave de partición es nulo, Amazon Redshift descarga de forma automática esos datos en una partición predeterminada que se conoce como `partition_column=__HIVE_DEFAULT_PARTITION__`.
- El comando UNLOAD no realiza ninguna llamada a un catálogo externo. Para registrar las nuevas particiones para que formen parte de la tabla externa existente, use un comando ALTER ... ADD PARTITION ... independiente. O puede ejecutar un comando CREATE EXTERNAL TABLE para registrar los datos descargados como una nueva tabla externa. También puede utilizar un rastreador de AWS Glue para completar el catálogo de datos. Para obtener más información, consulte Definición de rastreadores en la Guía para desarrolladores de AWS Glue.

- Si utiliza la opción MANIFEST, Amazon Redshift generará solo un archivo de manifiesto en la carpeta raíz de Amazon S3.
- Los tipos de datos de columna que puede utilizar como clave de partición son SMALLINT, INTEGER, BIGINT, DECIMAL, REAL, BOOLEAN, CHAR, VARCHAR, DATE y TIMESTAMP.

Uso del privilegio ASSUMEROLE para conceder acceso a un rol de IAM para operaciones UNLOAD

Para brindar acceso a usuarios y grupos específicos a un rol de IAM para operaciones UNLOAD, un superusuario puede conceder el privilegio ASSUMEROLE en un rol de IAM a dichos usuarios y grupos. Para obtener más información, consulta GRANT.

UNLOAD no admite alias de puntos de acceso de Amazon S3

Puede utilizar los alias de puntos de acceso de Amazon S3 con el comando UNLOAD.

## Ejemplos

Para ver ejemplos que muestran cómo utilizar el comando UNLOAD, consulte Ejemplos de UNLOAD.

## Ejemplos de UNLOAD

Estos ejemplos demuestran varios parámetros del comando UNLOAD. En muchos de los ejemplos se utilizan los datos de muestra de TICKIT. Para obtener más información, consulte Base de datos de muestra.

> **Note**
>
> Estos ejemplos contienen saltos de línea por motivos de legibilidad. No incluya saltos de líneas ni espacios en la cadena credentials-args (credenciales-argumentos).

Descargar VENUE en un archivo delimitado por la barra vertical (delimitador predeterminado)

En el siguiente ejemplo, se descarga la tabla VENUE y se escriben los datos en `s3://amzn-s3-demo-bucket/unload/`:

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/unload/'
```

```
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole';
```

Por defecto, UNLOAD escribe uno o más archivos por sector. Suponiendo que un clúster de dos nodos con dos sectores por nodo, el ejemplo anterior crea estos archivos en `amzn-s3-demo-bucket`:

```
unload/0000_part_00
unload/0001_part_00
unload/0002_part_00
unload/0003_part_00
```

Para diferenciar mejor los archivos de salida, puede incluir un prefijo en la ubicación. En el siguiente ejemplo, se descarga la tabla VENUE y se escriben los datos en `s3://amzn-s3-demo-bucket/unload/venue_pipe_`:

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/unload/venue_pipe_'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole';
```

El resultado es estos cuatro archivos en la carpeta `unload`, suponiendo nuevamente que hay cuatro secciones.

```
venue_pipe_0000_part_00
venue_pipe_0001_part_00
venue_pipe_0002_part_00
venue_pipe_0003_part_00
```

Descargar la tabla LINEITEM en archivos Parquet particionados

En el siguiente ejemplo, se descarga la tabla LINEITEM en formato Parquet, particionada por la columna `l_shipdate`.

```
unload ('select * from lineitem')
to 's3://amzn-s3-demo-bucket/lineitem/'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
PARQUET
PARTITION BY (l_shipdate);
```

Suponiendo que hay cuatro sectores, los archivos Parquet resultantes se dividen dinámicamente en varias carpetas.

```
s3://amzn-s3-demo-bucket/lineitem/l_shipdate=1992-01-02/0000_part_00.parquet
                                                0001_part_00.parquet
                                                0002_part_00.parquet
                                                0003_part_00.parquet
s3://amzn-s3-demo-bucket/lineitem/l_shipdate=1992-01-03/0000_part_00.parquet
                                                0001_part_00.parquet
                                                0002_part_00.parquet
                                                0003_part_00.parquet
s3://amzn-s3-demo-bucket/lineitem/l_shipdate=1992-01-04/0000_part_00.parquet
                                                0001_part_00.parquet
                                                0002_part_00.parquet
                                                0003_part_00.parquet
...
```

**Note**

En algunos casos, el comando UNLOAD utiliza la opción INCLUDE como se muestra en la siguiente instrucción SQL.

```
unload ('select * from lineitem')
to 's3://amzn-s3-demo-bucket/lineitem/'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
PARQUET
PARTITION BY (l_shipdate) INCLUDE;
```

En estos casos, la columna `l_shipdate` también está en los datos de los archivos Parquet. De lo contrario, los datos de la columna `l_shipdate` no están en los archivos Parquet.

Descargar la tabla VENUE en un archivo JSON

En el siguiente ejemplo, se descarga la tabla VENUE y se escriben los datos en formato JSON en `s3://amzn-s3-demo-bucket/unload/`.

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/unload/'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
JSON;
```

A continuación, se incluyen ejemplos de filas de la tabla VENUE.

```
venueid | venuename                  | venuecity       | venuestate | venueseats
--------+----------------------------+-----------------+------------+-----------
      1 | Pinewood Racetrack         | Akron           | OH         | 0
      2 | Columbus "Crew" Stadium    | Columbus        | OH         | 0
      4 | Community, Ballpark, Arena | Kansas City     | KS         | 0
```

Después de descargar en JSON, el formato del archivo es similar al siguiente.

```
{"venueid":1,"venuename":"Pinewood
 Racetrack","venuecity":"Akron","venuestate":"OH","venueseats":0}
{"venueid":2,"venuename":"Columbus \"Crew\" Stadium
 ","venuecity":"Columbus","venuestate":"OH","venueseats":0}
{"venueid":4,"venuename":"Community, Ballpark, Arena","venuecity":"Kansas
 City","venuestate":"KS","venueseats":0}
```

Descargar VENUE en un archivo CSV

En el siguiente ejemplo, se descarga la tabla VENUE y se escriben los datos en formato CSV en `s3://amzn-s3-demo-bucket/unload/`.

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/unload/'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
CSV;
```

Suponga que la tabla VENUE contiene las siguientes filas.

```
venueid | venuename                  | venuecity       | venuestate | venueseats
--------+----------------------------+-----------------+------------+-----------
      1 | Pinewood Racetrack         | Akron           | OH         | 0
      2 | Columbus "Crew" Stadium    | Columbus        | OH         | 0
      4 | Community, Ballpark, Arena | Kansas City     | KS         | 0
```

El archivo de descarga tiene un aspecto similar al siguiente.

```
1,Pinewood Racetrack,Akron,OH,0
2,"Columbus ""Crew"" Stadium",Columbus,OH,0
4,"Community, Ballpark, Arena",Kansas City,KS,0
```

Descargar VENUE en un archivo CSV mediante un delimitador

En el siguiente ejemplo, se descarga la tabla VENUE y se escriben los datos en formato CSV utilizando el carácter de barra vertical (|) como delimitador. El archivo descargado se escribe en `s3://amzn-s3-demo-bucket/unload/`. La tabla VENUE de este ejemplo contiene el carácter de barra vertical en el valor de la primera fila (`Pinewood Race|track`). Esto se hace para mostrar que el valor del resultado está entre comillas dobles. Las comillas dobles utilizan la secuencia de escape con una comilla doble y todo el campo se encierra entre comillas dobles.

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/unload/'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
CSV DELIMITER AS '|';
```

Suponga que la tabla VENUE contiene las siguientes filas.

```
venueid | venuename                  | venuecity       | venuestate | venueseats
--------+----------------------------+-----------------+------------+------------
      1 | Pinewood Race|track        | Akron           | OH         | 0
      2 | Columbus "Crew" Stadium    | Columbus        | OH         | 0
      4 | Community, Ballpark, Arena | Kansas City     | KS         | 0
```

El archivo de descarga tiene un aspecto similar al siguiente.

```
1|"Pinewood Race|track"|Akron|OH|0
2|"Columbus ""Crew"" Stadium"|Columbus|OH|0
4|Community, Ballpark, Arena|Kansas City|KS|0
```

Descargar VENUE con un archivo de manifiesto

Para crear un archivo de manifiesto, incluya la opción MANIFEST. En el siguiente ejemplo, se descarga la tabla VENUE y se escriben un archivo de manifiesto junto con los archivos de datos en s3://amzn-s3-demo-bucket/venue_pipe_:

> **Important**
>
> Si descarga archivos con la opción MANIFEST, debe utilizar la opción MANIFEST con el comando COPY cuando carga los archivos. Si utiliza el mismo prefijo para cargar los

archivos y no especifica la opción MANIFEST, COPY produce un error porque supone que el archivo de manifiesto es un archivo de datos.

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/venue_pipe_' iam_role 'arn:aws:iam::0123456789012:role/
MyRedshiftRole'
manifest;
```

El resultado es estos cinco archivos:

```
s3://amzn-s3-demo-bucket/venue_pipe_0000_part_00
s3://amzn-s3-demo-bucket/venue_pipe_0001_part_00
s3://amzn-s3-demo-bucket/venue_pipe_0002_part_00
s3://amzn-s3-demo-bucket/venue_pipe_0003_part_00
s3://amzn-s3-demo-bucket/venue_pipe_manifest
```

A continuación se muestra el contenido de un archivo de manifiesto.

```
{
  "entries": [
    {"url":"s3://amzn-s3-demo-bucket/tickit/venue_0000_part_00"},
    {"url":"s3://amzn-s3-demo-bucket/tickit/venue_0001_part_00"},
    {"url":"s3://amzn-s3-demo-bucket/tickit/venue_0002_part_00"},
    {"url":"s3://amzn-s3-demo-bucket/tickit/venue_0003_part_00"}
  ]
}
```

Descargar VENUE con MANIFEST VERBOSE

Cuando se especifica la opción MANIFEST VERBOSE, el archivo de manifiesto incluye las siguientes secciones:

- La sección `entries` enumera la ruta de Amazon S3, el tamaño del archivo y el recuento de filas de cada archivo.
- La sección `schema` enumera los nombres de columnas, los tipos de datos y la dimensión para cada columna.
- La sección `meta` muestra el tamaño del archivo total y el recuento de filas de todos los archivos.

En el siguiente ejemplo, se descarga la tabla VENUE usando la opción MANIFEST VERBOSE.

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/unload_venue_folder/'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
manifest verbose;
```

A continuación se muestra el contenido de un archivo de manifiesto.

```
{
  "entries": [
    {"url":"s3://amzn-s3-demo-bucket/venue_pipe_0000_part_00", "meta":
 { "content_length": 32295, "record_count": 10 }},
    {"url":"s3://amzn-s3-demo-bucket/venue_pipe_0001_part_00", "meta":
 { "content_length": 32771, "record_count": 20 }},
    {"url":"s3://amzn-s3-demo-bucket/venue_pipe_0002_part_00", "meta":
 { "content_length": 32302, "record_count": 10 }},
    {"url":"s3://amzn-s3-demo-bucket/venue_pipe_0003_part_00", "meta":
 { "content_length": 31810, "record_count": 15 }}
  ],
  "schema": {
    "elements": [
      {"name": "venueid", "type": { "base": "integer" }},
      {"name": "venuename", "type": { "base": "character varying", 25 }},
      {"name": "venuecity", "type": { "base": "character varying", 25 }},
      {"name": "venuestate", "type": { "base": "character varying", 25 }},
      {"name": "venueseats", "type": { "base": "character varying", 25 }}
    ]
  },
  "meta": {
    "content_length": 129178,
    "record_count": 55
  },
  "author": {
    "name": "Amazon Redshift",
    "version": "1.0.0"
  }
}
```

Descargar VENUE con un encabezado

En el siguiente ejemplo, se descarga VENUE con una fila de encabezado.

```
unload ('select * from venue where venueseats > 75000')
to 's3://amzn-s3-demo-bucket/unload/'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
header
parallel off;
```

A continuación, se muestra el contenido del archivo de salida con una fila de encabezado.

```
venueid|venuename|venuecity|venuestate|venueseats
6|New York Giants Stadium|East Rutherford|NJ|80242
78|INVESCO Field|Denver|CO|76125
83|FedExField|Landover|MD|91704
79|Arrowhead Stadium|Kansas City|MO|79451
```

Descargar VENUE en archivos más pequeños

Por defecto, el tamaño máximo de archivo es de 6,2 GB. Si los datos de descarga superan 6.2 GB, UNLOAD crea un nuevo archivo para cada segmento de datos de 6,2 GB. Para crear archivos más pequeños, incluya el parámetro MAXFILESIZE. Suponiendo que el tamaño de los datos en el ejemplo anterior era de 20 GB, el siguiente comando UNLOAD crea 20 archivos, cada uno de 1 GB.

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/unload/'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
maxfilesize 1 gb;
```

Descargar VENUE en serie

Para descargar en serie, especifique PARALLEL OFF. UNLOAD escribe un archivo a la vez, hasta un máximo de 6,2 GB por archivo.

En el siguiente ejemplo, se descarga la tabla VENUE y se escriben los datos en serie en `s3://amzn-s3-demo-bucket/unload/`.

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/unload/venue_serial_'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
parallel off;
```

El resultado es un archivo denominado venue_serial_000.

Si los datos de descarga superan 6.2 GB, UNLOAD crea un nuevo archivo para cada segmento de datos de 6,2 GB. En el siguiente ejemplo, se descarga la tabla LINEORDER y se escriben los datos en serie en `s3://amzn-s3-demo-bucket/unload/`.

```
unload ('select * from lineorder')
to 's3://amzn-s3-demo-bucket/unload/lineorder_serial_'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
parallel off gzip;
```

El resultado es la siguiente serie de archivos.

```
lineorder_serial_0000.gz
lineorder_serial_0001.gz
lineorder_serial_0002.gz
lineorder_serial_0003.gz
```

Para diferenciar mejor los archivos de salida, puede incluir un prefijo en la ubicación. En el siguiente ejemplo, se descarga la tabla VENUE y se escriben los datos en `s3://amzn-s3-demo-bucket/venue_pipe_`:

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/unload/venue_pipe_'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole';
```

El resultado es estos cuatro archivos en la carpeta `unload`, suponiendo nuevamente que hay cuatro secciones.

```
venue_pipe_0000_part_00
venue_pipe_0001_part_00
venue_pipe_0002_part_00
venue_pipe_0003_part_00
```

Cargar VENUE desde los archivos de descarga

Para cargar una tabla desde un conjunto de archivos de descarga, simplemente debe revertir el proceso al utilizar un comando COPY. En el siguiente ejemplo, se crea una nueva tabla, LOADVENUE, y se carga la tabla desde los archivos de datos creados en el ejemplo anterior.

```
create table loadvenue (like venue);
```

```
copy loadvenue from 's3://amzn-s3-demo-bucket/venue_pipe_' iam_role
 'arn:aws:iam::0123456789012:role/MyRedshiftRole';
```

Si utilizó la opción MANIFEST para crear un archivo de manifiesto con los archivos de descarga, puede cargar los datos a través del mismo archivo de manifiesto. Para ello, debe usar un comando COPY con la opción MANIFEST. En el siguiente ejemplo, se cargan datos a través de un archivo de manifiesto.

```
copy loadvenue
from 's3://amzn-s3-demo-bucket/venue_pipe_manifest' iam_role
 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
manifest;
```

Descargar VENUE en archivos cifrados

En el siguiente ejemplo, se descarga la tabla VENUE en un conjunto de archivos cifrados mediante una clave de AWS KMS. Si especifica un archivo de manifiesto con la opción ENCRYPTED, también se cifra el archivo de manifiesto. Para obtener más información, consulte Descarga de archivos de datos cifrados.

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/venue_encrypt_kms'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
kms_key_id '1234abcd-12ab-34cd-56ef-1234567890ab'
manifest
encrypted;
```

En el siguiente ejemplo, se descarga la tabla VENUE en un conjunto de archivos cifrados a través de una clave raíz simétrica.

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/venue_encrypt_cmk'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
master_symmetric_key 'EXAMPLEMASTERKEYtkbjk/OpCwtYSx/M4/t7DMCDIK722'
encrypted;
```

Cargar VENUE desde archivos cifrados

Para cargar tablas desde un conjunto de archivos que se crearon usando UNLOAD con la opción ENCRYPT, revierta el proceso utilizando un comando COPY. Con ese comando, use la opción

ENCRYPTED y especifique la misma clave raíz simétrica que se utilizó para el comando UNLOAD. En el siguiente ejemplo, se carga la tabla LOADVENUE desde los archivos de datos cifrados creados en el ejemplo anterior.

```
create table loadvenue (like venue);

copy loadvenue
from 's3://amzn-s3-demo-bucket/venue_encrypt_manifest'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
master_symmetric_key 'EXAMPLEMASTERKEYtkbjk/OpCwtYSx/M4/t7DMCDIK722'
manifest
encrypted;
```

Descargar datos de VENUE en un archivo delimitado por tabulaciones

```
unload ('select venueid, venuename, venueseats from venue')
to 's3://amzn-s3-demo-bucket/venue_tab_'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
delimiter as '\t';
```

Los archivos de datos de salida tendrán este aspecto:

```
1 Toyota Park Bridgeview IL 0
2 Columbus Crew Stadium Columbus OH 0
3 RFK Stadium Washington DC 0
4 CommunityAmerica Ballpark Kansas City KS 0
5 Gillette Stadium Foxborough MA 68756
...
```

Descargar VENUE en un archivo de datos de ancho fijo

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/venue_fw_'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
fixedwidth as 'venueid:3,venuename:39,venuecity:16,venuestate:2,venueseats:6';
```

Los archivos de datos de salida tendrán un aspecto similar al siguiente.

```
1  Toyota Park              Bridgeview  IL0
2  Columbus Crew Stadium    Columbus    OH0
3  RFK Stadium              Washington  DC0
```

```
4  CommunityAmerica BallparkKansas City KS0
5  Gillette Stadium         Foxborough  MA68756
...
```

Descargar VENUE en un conjunto de archivos comprimidos GZIP delimitados por tabulaciones

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/venue_tab_'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
delimiter as '\t'
gzip;
```

Descargar VENUE en un archivo de texto comprimido con GZIP

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/venue_tab_'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
extension 'txt.gz'
gzip;
```

Descargar datos que contienen un delimitador

Este ejemplo usa la opción ADDQUOTES para descargar datos delimitados por comas donde algunos campos de datos contienen una coma.

Primero, cree una tabla que contenga comillas.

```
create table location (id int, location char(64));

insert into location values (1,'Phoenix, AZ'),(2,'San Diego, CA'),(3,'Chicago, IL');
```

Luego, descargue los datos con la opción ADDQUOTES.

```
unload ('select id, location from location')
to 's3://amzn-s3-demo-bucket/location_'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
delimiter ',' addquotes;
```

Los archivos de datos descargados tendrán este aspecto:

```
1,"Phoenix, AZ"
2,"San Diego, CA"
```

```
3,"Chicago, IL"
...
```

Descargar los resultados de una consulta de combinación

En el siguiente ejemplo, se descargan los resultados de una consulta de combinación que contiene una función de ventana.

```
unload ('select venuecity, venuestate, caldate, pricepaid,
sum(pricepaid) over(partition by venuecity, venuestate
order by caldate rows between 3 preceding and 3 following) as winsum
from sales join date on sales.dateid=date.dateid
join event on event.eventid=sales.eventid
join venue on event.venueid=venue.venueid
order by 1,2')
to 's3://amzn-s3-demo-bucket/tickit/winsum'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole';
```

Los archivos de salida tendrán este aspecto:

```
Atlanta|GA|2008-01-04|363.00|1362.00
Atlanta|GA|2008-01-05|233.00|2030.00
Atlanta|GA|2008-01-06|310.00|3135.00
Atlanta|GA|2008-01-08|166.00|8338.00
Atlanta|GA|2008-01-11|268.00|7630.00
...
```

Descargar con NULL AS

Los valores nulos de resultado de UNLOAD son cadenas vacías por defecto. En los siguientes ejemplos se muestra cómo utilizar NULL AS para sustituir una cadena de texto para los valores nulos.

Para estos ejemplos, agregamos algunos valores nulos a la tabla VENUE.

```
update venue set venuestate = NULL
where venuecity = 'Cleveland';
```

Seleccione en VENUE donde VENUESTATE toma el valor nulo para verificar que las columnas contienen NULL.

```
select * from venue where venuestate is null;
```

```
 venueid |        venuename         | venuecity | venuestate | venueseats
---------+--------------------------+-----------+------------+------------
      22 | Quicken Loans Arena      | Cleveland |            |          0
     101 | Progressive Field        | Cleveland |            |      43345
      72 | Cleveland Browns Stadium | Cleveland |            |      73200
```

Ahora, aplique UNLOAD en la tabla VENUE con la opción NULL AS para reemplazar valores nulos con la cadena de caracteres "`fred`".

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/nulls/'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
null as 'fred';
```

En el siguiente ejemplo del archivo de descarga se muestra que los valores nulos se reemplazaron con `fred`. Resulta que algunos valores para VENUESEATS también eran valores nulos y se reemplazaron con `fred`. Aunque el tipo de datos para VENUESEATS es entero, el comando UNLOAD convierte los valores a texto en los archivos de descarga y, luego, el comando COPY los convierte nuevamente a valores enteros. Si está descargando un archivo de ancho fijo, la cadena NULL AS no debe ser más grande que el ancho del campo.

```
248|Charles Playhouse|Boston|MA|0
251|Paris Hotel|Las Vegas|NV|fred
258|Tropicana Hotel|Las Vegas|NV|fred
300|Kennedy Center Opera House|Washington|DC|0
306|Lyric Opera House|Baltimore|MD|0
308|Metropolitan Opera|New York City|NY|0
  5|Gillette Stadium|Foxborough|MA|5
 22|Quicken Loans Arena|Cleveland|fred|0
101|Progressive Field|Cleveland|fred|43345
...
```

Para cargar una tabla desde los archivos de descarga, utilice un comando COPY con la misma opción NULL AS.

> **Note**
>
> Si intenta cargar valores nulos en una columna definida como NOT NULL, el comando COPY produce un error.

```
create table loadvenuenulls (like venue);

copy loadvenuenulls from 's3://amzn-s3-demo-bucket/nulls/'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
null as 'fred';
```

Para verificar que las columnas contienen un valor nulo, no solo cadenas vacías, seleccione LOADVENUENULLS y filtre en busca de valores nulos.

```
select * from loadvenuenulls where venuestate is null or venueseats is null;

 venueid |        venuename         | venuecity | venuestate | venueseats
---------+--------------------------+-----------+------------+------------
      72 | Cleveland Browns Stadium | Cleveland |            |      73200
     253 | Mirage Hotel             | Las Vegas | NV         |
     255 | Venetian Hotel           | Las Vegas | NV         |
      22 | Quicken Loans Arena      | Cleveland |            |          0
     101 | Progressive Field        | Cleveland |            |      43345
     251 | Paris Hotel              | Las Vegas | NV         |

...
```

Puede utilizar UNLOAD en una tabla que contiene valores nulos con el comportamiento NULL AS predeterminado y, luego, aplicar COPY para copiar los datos nuevamente en una tabla a través del comportamiento NULL AS predeterminado. No obstante, los campos no numéricos en la tabla de destino tendrá cadenas vacías, no nulas. Por defecto, UNLOAD convierte valores nulos en cadenas vacías (espacio en blanco o longitud cero). COPY convierte cadenas vacías en valores NULL para columnas numéricas, pero inserta cadenas vacías en columnas no numéricas. En el siguiente ejemplo, se muestra cómo realizar un UNLOAD seguido de un comando COPY a través del comportamiento NULL AS predeterminado.

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/nulls/'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole' allowoverwrite;

truncate loadvenuenulls;
copy loadvenuenulls from 's3://amzn-s3-demo-bucket/nulls/'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole';
```

En este caso, cuando filtró en busca de valores nulos, solo las filas donde VENUESEATS contenía valores nulos. Donde VENUESTATE contenía valores nulos en la tabla (VENUE), VENUESTATE en la tabla de destino (LOADVENUENULLS) contenía cadenas vacías.

```
select * from loadvenuenulls where venuestate is null or venueseats is null;

 venueid |        venuename         | venuecity | venuestate | venueseats
---------+--------------------------+-----------+------------+------------
     253 | Mirage Hotel             | Las Vegas | NV         |
     255 | Venetian Hotel           | Las Vegas | NV         |
     251 | Paris Hotel              | Las Vegas | NV         |
...
```

Para cargar cadenas vacías en columnas no numéricas como NULL, incluya las opciones EMPTYASNULL o BLANKSASNULL. Se pueden utilizar ambas.

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/nulls/'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole' allowoverwrite;

truncate loadvenuenulls;
copy loadvenuenulls from 's3://amzn-s3-demo-bucket/nulls/'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole' EMPTYASNULL;
```

Para verificar que las columnas contienen un valor NULL, no solo espacios en blanco o cadenas vacías, seleccione LOADVENUENULLS y filtre el contenido en busca de valores nulos.

```
select * from loadvenuenulls where venuestate is null or venueseats is null;

 venueid |        venuename         | venuecity | venuestate | venueseats
---------+--------------------------+-----------+------------+------------
      72 | Cleveland Browns Stadium | Cleveland |            |      73200
     253 | Mirage Hotel             | Las Vegas | NV         |
     255 | Venetian Hotel           | Las Vegas | NV         |
      22 | Quicken Loans Arena      | Cleveland |            |          0
     101 | Progressive Field        | Cleveland |            |      43345
     251 | Paris Hotel              | Las Vegas | NV         |
     ...
```

Descargar mediante el parámetro ALLOWOVERWRITE

De manera predeterminada, UNLOAD no sobrescribe los archivos existentes en el bucket de destino. Por ejemplo, si ejecuta la misma instrucción UNLOAD dos veces sin modificar los archivos en el bucket de destino, el segundo comando UNLOAD produce un error. Para sobrescribir los archivos existentes, incluido el archivo de manifiesto, especifique la opción ALLOWOVERWRITE.

```
unload ('select * from venue')
to 's3://amzn-s3-demo-bucket/venue_pipe_'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole'
manifest allowoverwrite;
```

Descargar la tabla EVENT mediante los parámetros PARALLEL y MANIFEST

Puede usar UNLOAD para descargar una tabla en paralelo y generar un archivo de manifiesto. Los archivos de datos de Amazon S3 se crean todos al mismo nivel y los nombres tienen como sufijo el patrón `0000_part_00`. El archivo de manifiesto se encuentra en el mismo nivel de carpeta que los archivos de datos y tiene como sufijo el texto `manifest`. El siguiente código SQL descarga la tabla EVENT y crea archivos con el nombre base `parallel`

```
unload ('select * from mytickit1.event')
to 's3://amzn-s3-demo-bucket/parallel'
iam_role 'arn:aws:iam::123456789012:role/MyRedshiftRole'
parallel on
manifest;
```

La lista de archivos de Amazon S3 es similar a la siguiente.

```
 Name                      Last modified                         Size
 parallel0000_part_00 -   August 2, 2023, 14:54:39 (UTC-07:00) 52.1 KB
 parallel0001_part_00 -   August 2, 2023, 14:54:39 (UTC-07:00) 53.4 KB
 parallel0002_part_00 -   August 2, 2023, 14:54:39 (UTC-07:00) 52.1 KB
 parallel0003_part_00 -   August 2, 2023, 14:54:39 (UTC-07:00) 51.1 KB
 parallel0004_part_00 -   August 2, 2023, 14:54:39 (UTC-07:00) 54.6 KB
 parallel0005_part_00 -   August 2, 2023, 14:54:39 (UTC-07:00) 53.4 KB
 parallel0006_part_00 -   August 2, 2023, 14:54:39 (UTC-07:00) 54.1 KB
 parallel0007_part_00 -   August 2, 2023, 14:54:39 (UTC-07:00) 55.9 KB
 parallelmanifest       -   August 2, 2023, 14:54:39 (UTC-07:00) 886.0 B
```

El contenido del archivo `parallelmanifest` es similar al siguiente.

```
{
  "entries": [
    {"url":"s3://amzn-s3-demo-bucket/parallel0000_part_00", "meta": { "content_length":
 53316 }},
    {"url":"s3://amzn-s3-demo-bucket/parallel0001_part_00", "meta": { "content_length":
 54704 }},
    {"url":"s3://amzn-s3-demo-bucket/parallel0002_part_00", "meta": { "content_length":
 53326 }},
    {"url":"s3://amzn-s3-demo-bucket/parallel0003_part_00", "meta": { "content_length":
 52356 }},
    {"url":"s3://amzn-s3-demo-bucket/parallel0004_part_00", "meta": { "content_length":
 55933 }},
    {"url":"s3://amzn-s3-demo-bucket/parallel0005_part_00", "meta": { "content_length":
 54648 }},
    {"url":"s3://amzn-s3-demo-bucket/parallel0006_part_00", "meta": { "content_length":
 55436 }},
    {"url":"s3://amzn-s3-demo-bucket/parallel0007_part_00", "meta": { "content_length":
 57272 }}
  ]
}
```

Descargar la tabla EVENT mediante los parámetros PARALLEL OFF y MANIFEST

Puede usar UNLOAD para descargar una tabla en serie (PARALLEL OFF) y generar un archivo de manifiesto. Los archivos de datos de Amazon S3 se crean todos al mismo nivel y los nombres tienen como sufijo el patrón `0000`. El archivo de manifiesto se encuentra en el mismo nivel de carpeta que los archivos de datos y tiene como sufijo el texto `manifest`.

```
unload ('select * from mytickit1.event')
to 's3://amzn-s3-demo-bucket/serial'
iam_role 'arn:aws:iam::123456789012:role/MyRedshiftRole'
parallel off
manifest;
```

La lista de archivos de Amazon S3 es similar a la siguiente.

```
 Name                    Last modified                      Size
 serial0000          -   August 2, 2023, 15:54:39 (UTC-07:00) 426.7 KB
```

```
 serialmanifest         -   August 2, 2023, 15:54:39 (UTC-07:00) 120.0 B
```

El contenido del archivo `serialmanifest` es similar al siguiente.

```
{
  "entries": [
    {"url":"s3://amzn-s3-demo-bucket/serial000", "meta": { "content_length": 436991 }}
  ]
}
```

Descargar la tabla EVENT mediante los parámetros PARTITION BY y MANIFEST

Puede usar UNLOAD para descargar una tabla por partición y generar un archivo de manifiesto. Se crea una nueva carpeta en Amazon S3 con carpetas de partición secundarias y los archivos de datos en las carpetas secundarias con un patrón de nombres similar a `0000_par_00`. El archivo de manifiesto se encuentra en el mismo nivel de carpeta que las carpetas secundarias con el nombre `manifest`.

```
unload ('select * from mytickit1.event')
to 's3://amzn-s3-demo-bucket/partition'
iam_role 'arn:aws:iam::123456789012:role/MyRedshiftRole'
partition by (eventname)
manifest;
```

La lista de archivos de Amazon S3 es similar a la siguiente.

```
 Name                   Type     Last modified                     Size

 partition            Folder
```

En la carpeta `partition` están las carpetas secundarias con el nombre de la partición y el archivo de manifiesto. A continuación, se muestra la parte final de la lista de carpetas de la carpeta `partition`, similar a la siguiente.

```
 Name                   Type      Last modified                      Size

 ...
```

```
 eventname=Zucchero/    Folder
 eventname=Zumanity/    Folder
 eventname=ZZ Top/      Folder
 manifest               -     August 2, 2023, 15:54:39 (UTC-07:00) 467.6 KB
```

En la carpeta `eventname=Zucchero/` están los archivos de datos similares a los siguientes.

```
 Name               Last modified                         Size
 0000_part_00 -   August 2, 2023, 15:59:19 (UTC-07:00) 70.0 B
 0001_part_00 -   August 2, 2023, 15:59:16 (UTC-07:00) 106.0 B
 0002_part_00 -   August 2, 2023, 15:59:15 (UTC-07:00) 70.0 B
 0004_part_00 -   August 2, 2023, 15:59:17 (UTC-07:00) 141.0 B
 0006_part_00 -   August 2, 2023, 15:59:16 (UTC-07:00) 35.0 B
 0007_part_00 -   August 2, 2023, 15:59:19 (UTC-07:00) 108.0 B
```

La parte final del contenido del archivo `manifest` es similar a la siguiente.

```
{
  "entries": [
    ...
    {"url":"s3://amzn-s3-demo-bucket/partition/eventname=Zucchero/007_part_00", "meta":
 { "content_length": 108 }},
    {"url":"s3://amzn-s3-demo-bucket/partition/eventname=Zumanity/007_part_00", "meta":
 { "content_length": 72 }}
  ]
}
```

Descargar la tabla EVENT con los parámetros MAXFILESIZE, ROWGROUPSIZE y MANIFEST

Puede usar UNLOAD para descargar una tabla en paralelo y generar un archivo de manifiesto. Los archivos de datos de Amazon S3 se crean todos al mismo nivel y los nombres tienen como sufijo el patrón `0000_part_00`. Los archivos de datos Parquet generados están limitados a 256 MB y el tamaño del grupo de filas a 128 MB. El archivo de manifiesto se encuentra en el mismo nivel de carpeta que los archivos de datos y tiene como sufijo `manifest`.

```
unload ('select * from mytickit1.event')
to 's3://amzn-s3-demo-bucket/eventsize'
iam_role 'arn:aws:iam::123456789012:role/MyRedshiftRole'
```

```
maxfilesize 256 MB
rowgroupsize 128 MB
parallel on
parquet
manifest;
```

La lista de archivos de Amazon S3 es similar a la siguiente.

```
 Name                            Type      Last modified                         Size
 eventsize0000_part_00.parquet parquet August 2, 2023, 17:35:21 (UTC-07:00) 24.5 KB
 eventsize0001_part_00.parquet parquet August 2, 2023, 17:35:21 (UTC-07:00) 24.8 KB
 eventsize0002_part_00.parquet parquet August 2, 2023, 17:35:21 (UTC-07:00) 24.4 KB
 eventsize0003_part_00.parquet parquet August 2, 2023, 17:35:21 (UTC-07:00) 24.0 KB
 eventsize0004_part_00.parquet parquet August 2, 2023, 17:35:21 (UTC-07:00) 25.3 KB
 eventsize0005_part_00.parquet parquet August 2, 2023, 17:35:21 (UTC-07:00) 24.8 KB
 eventsize0006_part_00.parquet parquet August 2, 2023, 17:35:21 (UTC-07:00) 25.0 KB
 eventsize0007_part_00.parquet parquet August 2, 2023, 17:35:21 (UTC-07:00) 25.6 KB
 eventsizemanifest                 -       August 2, 2023, 17:35:21 (UTC-07:00) 958.0 B
```

El contenido del archivo `eventsizemanifest` es similar al siguiente.

```
{
  "entries": [
    {"url":"s3://amzn-s3-demo-bucket/eventsize0000_part_00.parquet", "meta":
 { "content_length": 25130 }},
    {"url":"s3://amzn-s3-demo-bucket/eventsize0001_part_00.parquet", "meta":
 { "content_length": 25428 }},
    {"url":"s3://amzn-s3-demo-bucket/eventsize0002_part_00.parquet", "meta":
 { "content_length": 25025 }},
    {"url":"s3://amzn-s3-demo-bucket/eventsize0003_part_00.parquet", "meta":
 { "content_length": 24554 }},
    {"url":"s3://amzn-s3-demo-bucket/eventsize0004_part_00.parquet", "meta":
 { "content_length": 25918 }},
    {"url":"s3://amzn-s3-demo-bucket/eventsize0005_part_00.parquet", "meta":
 { "content_length": 25362 }},
    {"url":"s3://amzn-s3-demo-bucket/eventsize0006_part_00.parquet", "meta":
 { "content_length": 25647 }},
    {"url":"s3://amzn-s3-demo-bucket/eventsize0007_part_00.parquet", "meta":
 { "content_length": 26256 }}
  ]
```

```
}
```

# UPDATE

Temas

- Sintaxis
- Parámetros
- Notas de uso
- Ejemplos de instrucciones UPDATE

Actualiza valores en una o más columnas de la tabla cuando se cumple una condición.

> **Note**
>
> El tamaño máximo de una instrucción SQL es de 16 MB.

## Sintaxis

```
[ WITH [RECURSIVE] common_table_expression [, common_table_expression , ...] ]
            UPDATE table_name [ [ AS ] alias ] SET column = { expression | DEFAULT }
 [,...]

[ FROM fromlist ]
[ WHERE condition ]
```

## Parámetros

Cláusula WITH

Se trata de una cláusula opcional que especifica una o más common-table-expressions. Consulte Cláusula WITH.

table_name

Una tabla temporal o persistente. Solo el propietario de la tabla o un usuario con el privilegio UPDATE en la tabla puede actualizar filas. Si utiliza la cláusula FROM o selecciona desde tablas en una expresión o condición, debe tener el privilegio SELECT en esas tablas. No puede asignar un alias a la tabla aquí; no obstante, puede especificar un alias en la cláusula FROM.

> **Note**
>
> Las tablas externas de Amazon Redshift Spectrum solo se pueden leer. No puede utilizar el comando UPDATE en una tabla externa.

alias

Se trata del nombre alternativo temporal para una tabla de destino. Los alias son opcionales. La palabra clave AS es siempre opcional.

SET column =

Una o más columnas que desea modificar. Las columnas que no están en la lista conservan sus valores actuales. No incluya el nombre de la tabla en la especificación de una columna de destino. Por ejemplo, `UPDATE tab SET tab.col = 1` no es válido.

expresión

Una expresión que define el nuevo valor para la columna especificada.

DEFAULT

Actualiza la columna con el valor predeterminado que se asignó a la columna en la instrucción CREATE TABLE.

FROM tablelist (lista_de_tablas)

Puede actualizar una tabla al hacer referencia a información de otras tablas. Enumere estas otras tablas en la cláusula FROM o use una subconsulta como parte de la condición WHERE. Las tablas enumeradas en la cláusula FROM pueden tener alias. Si necesita incluir la tabla de destino de la instrucción UPDATE en la lista, utilice un alias.

WHERE condition

La cláusula opcional que limita las actualizaciones a filas que coinciden con una condición. Cuando la condición devuelve `true`, se actualizan las columnas SET especificadas. La condición puede ser un predicado simple en una columna o una condición basada en el resultado de una subconsulta.

Puede otorgar un nombre a cualquier tabla de la subconsulta, incluida la tabla de destino para UPDATE.

## Notas de uso

Después de actualizar una gran cantidad de filas en una tabla:

- Limpie la tabla para recuperar espacio de almacenamiento y reordene las filas.
- Analice la tabla para actualizar las estadísticas para el planificador de consultas.

Las combinaciones externas izquierdas, derechas y completas no se admiten en la cláusula FROM de una instrucción UPDATE; devuelven el siguiente error:

```
ERROR: Target table must be part of an equijoin predicate
```

Si necesita especificar una combinación externa, utilice una subconsulta en la cláusula WHERE de la instrucción UPDATE.

Si la instrucción UPDATE requiere una combinación automática con la tabla de destino, necesita especificar la condición de la combinación, así como los criterios de la cláusula WHERE que califican filas para la operación de actualización. En general, cuando la tabla de destino se combina consigo misma o con otra tabla, una práctica recomendada es utilizar una subconsulta que separe claramente las condiciones de combinación de los criterios que califican filas para actualizaciones.

Las consultas UPDATE con varias coincidencias por fila arrojan un error cuando el parámetro de configuración `error_on_nondeterministic_update` se establece en true. Para obtener más información, consulte error_on_nondeterministic_update.

Puede actualizar una columna GENERATED BY DEFAULT AS IDENTITY. Las columnas definidas como GENERATED BY DEFAULT AS IDENTITY pueden actualizarse con los valores que se proporcionen. Para obtener más información, consulte GENERATED BY DEFAULT AS IDENTITY.

## Ejemplos de instrucciones UPDATE

Para obtener más información acerca de las tablas utilizadas en los ejemplos siguientes, consulte Base de datos de muestra.

La tabla CATEGORY de la base de datos TICKIT contiene las siguientes filas:

```
+-------+----------+-----------+--------------------------------------------+
| catid | catgroup |  catname  |                  catdesc                   |
+-------+----------+-----------+--------------------------------------------+
| 5     | Sports   | MLS       | Major League Soccer                        |
```

```
| 11    | Concerts | Classical | All symphony, concerto, and choir concerts |
| 1     | Sports   | MLB       | Major League Baseball                      |
| 6     | Shows    | Musicals  | Musical theatre                            |
| 3     | Sports   | NFL       | National Football League                   |
| 8     | Shows    | Opera     | All opera and light opera                  |
| 2     | Sports   | NHL       | National Hockey League                     |
| 9     | Concerts | Pop       | All rock and pop music concerts            |
| 4     | Sports   | NBA       | National Basketball Association            |
| 7     | Shows    | Plays     | All non-musical theatre                    |
| 10    | Concerts | Jazz      | All jazz singers and bands                 |
+-------+----------+-----------+--------------------------------------------+
```

Actualización de una tabla en función de un rango de valores

Actualice la columna CATGROUP en función de un rango de valores en la columna CATID.

```
UPDATE category
SET catgroup='Theatre'
WHERE catid BETWEEN 6 AND 8;

SELECT * FROM category
WHERE catid BETWEEN 6 AND 8;

+-------+----------+----------+---------------------------+
| catid | catgroup | catname  |          catdesc          |
+-------+----------+----------+---------------------------+
| 6     | Theatre  | Musicals | Musical theatre           |
| 7     | Theatre  | Plays    | All non-musical theatre   |
| 8     | Theatre  | Opera    | All opera and light opera |
+-------+----------+----------+---------------------------+
```

Actualización de una tabla en función de un valor actual

Actualice las columnas CATNAME y CATDESC en función del valor de CATGROUP actual:

```
UPDATE category
SET catdesc=default, catname='Shows'
WHERE catgroup='Theatre';

SELECT * FROM category
WHERE catname='Shows';

+-------+----------+---------+---------+
```

```
| catid | catgroup | catname | catdesc |
+-------+----------+---------+---------+
| 6     | Theatre  | Shows   | NULL    |
| 7     | Theatre  | Shows   | NULL    |
| 8     | Theatre  | Shows   | NULL    |
+-------+----------+---------+---------+)
```

En este caso, la columna CATDESC se configuró como nula porque no se definió un valor predeterminado cuando se creó la tabla.

Ejecute los siguientes comandos para establecer los datos de la tabla CATEGORY de regreso a los valores originales:

```
TRUNCATE category;

COPY category
FROM 's3://redshift-downloads/tickit/category_pipe.txt'
DELIMITER '|'
IGNOREHEADER 1
REGION 'us-east-1'
IAM_ROLE default;
```

Actualización de una tabla en función del resultado de una subconsulta de la cláusula WHERE

Actualice la tabla CATEGORY en función del resultado de una subconsulta en la cláusula WHERE:

```
UPDATE category
SET catdesc='Broadway Musical'
WHERE category.catid IN
(SELECT category.catid FROM category
JOIN event ON category.catid = event.catid
JOIN venue ON venue.venueid = event.venueid
JOIN sales ON sales.eventid = event.eventid
WHERE venuecity='New York City' AND catname='Musicals');
```

Consulte la tabla actualizada:

```
SELECT * FROM category ORDER BY catid;

+-------+----------+-----------+--------------------------------------------+
| catid | catgroup |  catname  |                  catdesc                   |
+-------+----------+-----------+--------------------------------------------+
```

```
| 2     | Sports   | NHL       | National Hockey League                      |
| 3     | Sports   | NFL       | National Football League                    |
| 4     | Sports   | NBA       | National Basketball Association             |
| 5     | Sports   | MLS       | Major League Soccer                         |
| 6     | Shows    | Musicals  | Broadway Musical                            |
| 7     | Shows    | Plays     | All non-musical theatre                     |
| 8     | Shows    | Opera     | All opera and light opera                   |
| 9     | Concerts | Pop       | All rock and pop music concerts             |
| 10    | Concerts | Jazz      | All jazz singers and bands                  |
| 11    | Concerts | Classical | All symphony, concerto, and choir concerts |
+-------+----------+-----------+---------------------------------------------+
```

Actualización de una tabla en función del resultado de una subconsulta de la cláusula WITH

Para actualizar la tabla CATEGORY basándose en el resultado de una subconsulta utilizando la cláusula WITH, utilice el siguiente ejemplo.

```
WITH u1 as (SELECT catid FROM event ORDER BY catid DESC LIMIT 1)
UPDATE category SET catid='200' FROM u1 WHERE u1.catid=category.catid;

SELECT * FROM category ORDER BY catid DESC LIMIT 1;

+-------+----------+---------+---------------------------------+
| catid | catgroup | catname |             catdesc             |
+-------+----------+---------+---------------------------------+
| 200   | Concerts | Pop     | All rock and pop music concerts |
+-------+----------+---------+---------------------------------+
```

Actualización de una tabla en función del resultado de una condición de combinación

Actualice las 11 filas originales en la tabla CATEGORY en función de las filas CATID coincidentes en la tabla EVENT:

```
UPDATE category SET catid=100
FROM event
WHERE event.catid=category.catid;

SELECT * FROM category ORDER BY catid;

+-------+----------+-----------+---------------------------------------------+
| catid | catgroup |  catname  |                 catdesc                     |
+-------+----------+-----------+---------------------------------------------+
| 2     | Sports   | NHL       | National Hockey League                      |
```

```
| 3     | Sports   | NFL       | National Football League                    |
| 4     | Sports   | NBA       | National Basketball Association             |
| 5     | Sports   | MLS       | Major League Soccer                         |
| 10    | Concerts | Jazz      | All jazz singers and bands                  |
| 11    | Concerts | Classical | All symphony, concerto, and choir concerts |
| 100   | Concerts | Pop       | All rock and pop music concerts             |
| 100   | Shows    | Plays     | All non-musical theatre                     |
| 100   | Shows    | Opera     | All opera and light opera                   |
| 100   | Shows    | Musicals  | Broadway Musical                            |
+-------+----------+-----------+---------------------------------------------+
```

Tenga en cuenta que la tabla EVENT está enumerada en la cláusula FROM y la condición de combinación de la tabla de destino está definida en la cláusula WHERE. Solo cuatro filas califican para la actualización. Estas cuatro filas son las filas cuyos valores CATID eran originalmente 6, 7, 8 y 9; solo estas cuatro categorías están representadas en la tabla EVENT:

```
SELECT DISTINCT catid FROM event;

+-------+
| catid |
+-------+
| 6     |
| 7     |
| 8     |
| 9     |
+-------+
```

Actualice las 11 filas originales en la tabla CATEGORY al extender el ejemplo anterior y agregar otra condición a la cláusula WHERE. Debido a la restricción en la columna CATGROUP, solo una fila califica para la actualización (aunque cuatro filas califican para la combinación).

```
UPDATE category SET catid=100
FROM event
WHERE event.catid=category.catid
AND catgroup='Concerts';

SELECT * FROM category WHERE catid=100;

+-------+----------+---------+---------------------------------+
| catid | catgroup | catname |             catdesc             |
+-------+----------+---------+---------------------------------+
| 100   | Concerts | Pop     | All rock and pop music concerts |
```

```
+-------+----------+---------+---------------------------------+
```

A continuación se muestra una manera alternativa de escribir este ejemplo:

```
UPDATE category SET catid=100
FROM event JOIN category cat ON event.catid=cat.catid
WHERE cat.catgroup='Concerts';
```

La ventaja de este enfoque es que los criterios de combinación están separados claramente de cualquier otro criterio que califica a las filas para la actualización. Observe el uso del alias CAT para la tabla CATEGORY en la cláusula FROM.

Actualizaciones con combinaciones externas en la cláusula FROM

El ejemplo anterior mostró una combinación interna especificada en la cláusula FROM de una instrucción UPDATE. El siguiente ejemplo devuelve un error, ya que la cláusula FROM no admite combinaciones externas en la tabla de destino:

```
UPDATE category SET catid=100
FROM event LEFT JOIN category cat ON event.catid=cat.catid
WHERE cat.catgroup='Concerts';
ERROR:  Target table must be part of an equijoin predicate
```

Si la combinación externa se requiere para la instrucción UPDATE, puede trasladar la sintaxis de la combinación interna a una subconsulta:

```
UPDATE category SET catid=100
FROM
(SELECT event.catid FROM event LEFT JOIN category cat ON event.catid=cat.catid)
 eventcat
WHERE category.catid=eventcat.catid
AND catgroup='Concerts';
```

Actualizaciones con columnas de otra tabla en la cláusula SET

Para actualizar la tabla listing de la base de datos de muestra TICKIT con los valores de la tabla sales, utilice el siguiente ejemplo.

```
SELECT listid, numtickets FROM listing WHERE sellerid = 1 ORDER BY 1 ASC LIMIT 5;

+--------+------------+
```

```
| listid | numtickets |
+--------+------------+
| 100423 | 4          |
| 108334 | 24         |
| 117150 | 4          |
| 135915 | 20         |
| 205927 | 6          |
+--------+------------+

UPDATE listing
SET numtickets = sales.sellerid
FROM sales
WHERE sales.sellerid = 1 AND listing.sellerid = sales.sellerid;

SELECT listid, numtickets FROM listing WHERE sellerid = 1 ORDER BY 1 ASC LIMIT 5;

+--------+------------+
| listid | numtickets |
+--------+------------+
| 100423 | 1          |
| 108334 | 1          |
| 117150 | 1          |
| 135915 | 1          |
| 205927 | 1          |
+--------+------------+
```

# USE

Cambia la base de datos en la que se ejecutan las consultas. SHOW USE apunta a la base de datos que se utilizó más recientemente con el comando USE. RESET USE restablece la base de datos utilizada. Esto significa que si la base de datos no se especifica en SQL, los objetos se buscan en la base de datos actual.

## Sintaxis

```
USE database
```

## Ejemplos

Supongamos que hay tres bases de datos, `dev` y `pdb`, y `pdb2`. Supongamos que hay dos tablas `t` en los esquemas públicos de cada una de las bases de datos. En primer lugar, inserte los datos en tablas de diferentes bases de datos:

```
dev=# insert into dev.public.t values (1);
INSERT 0 1
dev=# insert into pdb.public.t values (2);
INSERT 0 1
```

Sin configurar una base de datos de forma explícita, el sistema utiliza la base de datos conectada. Compruebe el contexto actual de la base de datos:

```
dev=# show use;
Use Database

(1 row)
dev=> show search_path;
search_path
$user, public
(1 row)
```

Al consultar una tabla `t` sin especificar una base de datos, el sistema utiliza la tabla de la base de datos actual:

```
dev=# select * from t;
c
----
1
(1 row)
```

Utilice el comando `use` para cambiar de bases de datos sin cambiar la conexión:

```
dev=# use pdb;
USE
dev=# show use;
 Use Database
--------------
 pdb
(1 row)
dev=# select * from t;
id
----
2
(1 row)
```

También puede especificar explícitamente el esquema:

```
dev=# select * from public.t;
id
----
2
(1 row)
```

Ahora puede crear tablas en diferentes esquemas dentro de la base de datos actual:

```
dev=# create table s1.t(id int);
CREATE TABLE
dev=# insert into pdb.s1.t values (3);
INSERT 0 1
```

La ruta de búsqueda determina a qué objetos del esquema se accede cuando no se especifica un esquema:

```
dev=# set search_path to public, s1;
SET
dev=# select * from t;
 id
----
  2
(1 row)
```

Cambie el orden de los esquemas para acceder a diferentes tablas:

```
dev=# set search_path to s1, public;
SET
dev=# show search_path;
 search_path
-------------
 s1, public
(1 row)
dev=# select * from t;
 id
----
  3
(1 row)
```

Cambie a otra base de datos manteniendo su conexión original:

```
dev=# show use;
 Use Database
--------------
 pdb
(1 row)
dev=# use pdb2;
USE
dev=# show use;
 Use Database
--------------
 pdb2
(1 row)
```

Al cambiar de base de datos, la ruta de búsqueda se restablece a la predeterminada:

```
dev=# show search_path;
  search_path
---------------
 $user, public
(1 row)
```

Cree una tabla e inserte datos en la base de datos actual:

```
dev=# create table pdb2.public.t(id int);
CREATE TABLE
dev=# insert into pdb2.public.t values (4);
INSERT 0 1
dev=# select * from t;
 id
----
  4
(1 row)
```

En las transacciones, puede escribir en la base de datos actual y leer desde cualquier base de datos mediante una notación de tres partes. Esto también incluye la base de datos conectada:

```
dev=# show use;
 Use Database
--------------
 pdb2
(1 row)
```

```
dev=# BEGIN;
BEGIN
dev=# select * from t;
 id
----
  4
(1 row)

dev=# insert into t values (5);
INSERT 0 1
dev=# select * from t;
 id
----
  4
  5
(2 rows)

dev=# select * from pdb.public.t;
 id
----
  2
(1 row)

dev=# select * from dev.public.t;
 id
----
  1
(1 row)
```

Restablezca la base de datos conectada. Tenga en cuenta que esto no solo revierte a la base de datos pdb utilizada anteriormente, sino que se restablece a la base de datos conectada. La ruta de búsqueda también cambia a la predeterminada:

```
dev=# RESET USE;
RESET
dev=# select * from t;
c
----
1
(1 row)
dev=# show use;
 Use Database
```

```
--------------

(1 row)

dev=# show search_path;
  search_path
---------------
 $user, public
(1 row)
```

Puede cambiar las bases de datos al inicio de una transacción, pero no después de ejecutar las consultas:

```
dev=# BEGIN;
BEGIN
dev=# use pdb;
USE
dev=# use pdb2;
USE
dev=# use pdb;
USE
dev=# select * from t;
 id
----
  2
(1 row)
dev=# use pdb2;
ERROR:  USEd Database cannot be set or reset inside a transaction after another
 command.
dev=# rollback;
ROLLBACK
(1 row)
```

Ejemplo del catálogo de datos

En primer lugar, cree tablas en diferentes esquemas y catálogos para demostrar las consultas entre catálogos. Comience por crear tablas en la base de datos conectada.

```

dev=# CREATE TABLE dev.public.t (col INT);
dev=# INSERT INTO dev.public.t VALUES (1);
dev=# CREATE SCHEMA write_schema;
dev=# CREATE TABLE dev.write_schema.t (state char (2));
```

```
dev=# INSERT INTO dev.write_schema.t VALUES ('WA');
```

Ahora, cree tablas similares en un catálogo diferente. Esto demuestra cómo trabajar con bases de datos entre catálogos.

```
dev=# CREATE TABLE my_db@my_catalog.public.t (col INT);
dev=# INSERT INTO my_db@my_catalog.public.t VALUES (100);
dev=# CREATE SCHEMA my_db@my_catalog.write_schema;
dev=# CREATE TABLE my_db@my_catalog.write_schema.t (state char (2));
dev=# INSERT INTO my_db@my_catalog.write_schema.t VALUES ('CA');
```

Compruebe el contexto actual de la base de datos. Sin configurar una base de datos de forma explícita, el sistema utiliza la base de datos conectada.

```
dev=# SHOW USE;
 Use Database
--------------

(1 row)

dev=# SHOW search_path;
  search_path
---------------
 $user, public
(1 row)

dev=# SELECT * FROM t;
 col
-----
   1
(1 row)
```

Configure la base de datos USEd para consultar las tablas en un catálogo diferente.

```
dev=# USE my_db@my_catalog;
```

```
dev=# SHOW USE;
            Use Database
-------------------------------------
 my_db@my_catalog
(1 row)

dev=# SHOW search_path;
  search_path
---------------
 $user, public
(1 row)
```

Al consultar la tabla t, los resultados provienen de la base de datos entre catálogos.

```
dev=# SELECT * FROM t;
 col
-----
 100
(1 row)

dev=# SELECT * FROM public.t;
 col
-----
 100
(1 row)

dev=# SELECT * FROM my_db@my_catalog.public.t;
 col
-----
 100
(1 row)
```

Cambie la ruta de búsqueda para acceder a las tablas de distintos esquemas dentro de la base de datos USEd.

```
dev=# SET search_path to write_schema;

dev=# SHOW search_path;
 search_path
--------------
```

```
 write_schema
(1 row)

dev=# SELECT * FROM t;
 state
-------
 CA
(1 row)

dev=# SELECT * FROM write_schema.t;
 state
-------
 CA
(1 row)

dev=# SELECT * FROM my_db@my_catalog.write_schema.t;
 state
-------
 CA
(1 row)
```

Aunque USE está configurado como una base de datos entre catálogos, aún es posible realizar consultas explícitas en la base de datos original.

```
dev=# SELECT * FROM dev.write_schema.t;
 state
-------
 WA
(1 row)
```

Restablezca la base de datos USEd para que vuelva a hacer referencia a los objetos de la base de datos conectada.

```
dev=# RESET USE;

dev=# SHOW USE;
 Use Database
--------------
```

```
(1 row)
```

Tenga en cuenta que search_path se restablece cuando se restablece USE.

```
dev=# SHOW search_path;
  search_path
---------------
 $user, public
(1 row)
```

Tras el restablecimiento, las consultas ahora hacen referencia a la base de datos conectada original.

```
dev=# SELECT * FROM t;
 col
-----
   1
(1 row)

dev=# SELECT * FROM public.t;
 col
-----
   1
(1 row)

dev=# SELECT * FROM dev.public.t;
 col
-----
   1
(1 row)
```

Puede modificar la ruta de búsqueda en la base de datos original para acceder a diferentes esquemas.

```
dev=# SET search_path to write_schema;
```

```
dev=# SHOW search_path;
 search_path
--------------
 write_schema
(1 row)

dev=# SELECT * FROM t;
 state
-------
 WA
(1 row)

dev=# SELECT * FROM write_schema.t;
 state
-------
 WA
(1 row)

dev=# SELECT * FROM dev.write_schema.t;
 state
-------
 WA
(1 row)
```

## VACUUM

Reordena las filas y recupera espacio en una tabla especificada o en todas las tablas de la base de datos actual.

> **Note**
>
> Solo los usuarios con los permisos de tabla necesarios pueden vaciar una tabla de forma eficaz. Si se ejecuta VACUUM sin los privilegios de tabla necesarios, la operación se completa correctamente pero no tiene ningún efecto. Para obtener una lista de los permisos de tabla válidos para ejecutar VACUUM de forma eficaz, consulte la siguiente sección de privilegios requeridos.

Amazon Redshift ordena los datos de forma automática y ejecuta VACUUM DELETE en segundo plano. Esto disminuye la necesidad de ejecutar el comando VACUUM. Para obtener más información, consulte Limpieza de tablas.

De manera predeterminada, VACUUM omite la fase de ordenación para cualquier tabla que tenga más del 95 por ciento de las filas de la tabla ordenadas. La omisión de la fase de ordenación puede mejorar significativamente el rendimiento de VACUUM. Para cambiar el orden predeterminado o eliminar el umbral de una tabla, incluya el nombre de la tabla y el parámetro TO threshold (umbral) PERCENT cuando ejecute VACUUM.

Los usuarios pueden obtener acceso a las tablas mientras se limpian. Puede realizar consultas y escribir operaciones mientras se limpia una tabla, pero cuando se ejecutan comandos de lenguaje de manipulación de datos (DML) y una limpieza de manera simultánea, ambas operaciones podrían demorar más de lo usual. Si se ejecutan las instrucciones UPDATE y DELETE durante una limpieza, el rendimiento del sistema podría reducirse. VACUUM DELETE bloquea temporalmente las operaciones UPDATE y DELETE.

Amazon Redshift realiza automáticamente una operación de limpieza DELETE ONLY en segundo plano. La operación automática de limpieza se detiene cuando los usuarios ejecutan operaciones de lenguaje de definición de datos (DDL), tales como ALTER TABLE.

> **Note**
>
> La sintaxis y el comportamiento del comando VACUUM de Amazon Redshift son considerablemente diferentes de la operación VACUUM de PostgreSQL. Por ejemplo, la operación VACUUM predeterminada en Amazon Redshift es VACUUM FULL, que recupera espacio en el disco y reordena todas las filas. En cambio, la operación VACUUM predeterminada en PostgreSQL solo recupera espacio y lo pone a disposición para volver a utilizarlo.

Para obtener más información, consulte Limpieza de tablas.

## Privilegios necesarios

Los siguientes privilegios son necesarios para VACUUM:

- Superusuario
- Usuarios con el privilegio VACUUM
- Propietario de la tabla
- Propietario de la base de datos con quien se comparte la tabla

## Sintaxis

```
VACUUM [ FULL | SORT ONLY | DELETE ONLY | REINDEX | RECLUSTER ]
[ [ table_name ] [ TO threshold PERCENT ] [ BOOST ] ]
```

## Parámetros

FULL

Ordena la tabla especificada (o todas las tablas en la base de datos actual) y recupera espacio en disco ocupado por filas que las operaciones UPDATE y DELETE anteriores habían marcado para eliminación. El valor predeterminado es VACUUM FULL.

Una limpieza completa no realiza una reindexación para tablas intercaladas. Para reindexar tablas intercaladas seguido por una limpieza completa, utilice la opción VACUUM REINDEX.

Por defecto, VACUUM FULL omite la fase de ordenación de cualquier tabla que ya tenga al menos 95 por ciento ordenado. Si VACUUM puede omitir la fase de ordenación, ejecuta un comando DELETE ONLY y recupera espacio en la fase de eliminación para que al menos el 95 por ciento de las filas restantes no se marque para su eliminación.

Si no se cumple el umbral de ordenación (por ejemplo, si el 90 por ciento de las filas está ordenado) y VACUUM realiza una ordenación completa, también llevará a cabo una operación de eliminación completa y recuperará el espacio del 100 por cien de las filas eliminadas.

Puede cambiar el umbral de limpieza predeterminado solo para una tabla. Para cambiar el umbral de limpieza predeterminado de una sola tabla, incluya el nombre de la tabla y el parámetro TO threshold (umbral) PERCENT.

SORT ONLY

Ordena la tabla especificada (o todas las tablas de la base de datos actual) sin recuperar espacio liberado por filas eliminadas. Esta opción es útil cuando no es importante recuperar espacio en el disco, pero sí lo es la reordenación de nuevas filas. Una limpieza SORT ONLY reduce el tiempo transcurrido para las operaciones de limpieza cuando la región sin ordenar no contiene una gran cantidad de filas eliminadas o no abarca la región ordenada completa. Este tipo de limpieza puede ser beneficioso para las aplicaciones que no tienen limitaciones de espacio en el disco pero dependen de optimizaciones de consultas asociadas con mantener las filas de la tabla ordenadas.

Por defecto, VACUUM SORT ONLY omite cualquier tabla que ya tenga al menos 95 por ciento ordenado. Para cambiar el umbral de orden predeterminado de una tabla, incluya el nombre de la tabla y el parámetro TO threshold PERCENT cuando ejecute VACUUM.

DELETE ONLY

Amazon Redshift realiza de forma automática una operación de limpieza DELETE ONLY en segundo plano; por lo tanto, rara vez tendrá que ejecutar una limpieza DELETE ONLY.

Una VACUUM DELETE recupera espacio en disco ocupado por filas que las operaciones UPDATE y DELETE anteriores habían marcado para eliminación y compacta la tabla para liberar el espacio consumido. Una operación de limpieza DELETE ONLY no ordena los datos de la tabla.

Esta opción reduce el tiempo transcurrido en las operaciones de limpieza cuando la recuperación de espacio en el disco es importante, pero no la reordenación de nuevas filas. Esta opción también puede ser útil cuando el rendimiento de la consulta ya es óptimo y no es necesario reordenar las filas para optimizar el rendimiento de la consulta.

Por defecto, VACUUM DELETE ONLY recupera espacio para que al menos el 95 por ciento de las filas restantes no se marque para su eliminación. Para cambiar el umbral de eliminación predeterminado de una tabla, incluya el nombre de la tabla y el parámetro TO threshold PERCENT cuando ejecute VACUUM.

Algunas operaciones, como `ALTER TABLE APPEND`, pueden hacer que se fragmenten las tablas. Cuando se usa la cláusula `DELETE ONLY`, la operación vacuum reclama espacio de las tablas fragmentadas. El mismo valor de umbral del 95 por ciento se aplica a la operación de desfragmentación.

REINDEX

Analiza la distribución de los valores en columnas de claves de ordenación intercaladas y, luego, realiza una operación VACUUM completa. Si se utiliza REINDEX, se requiere un nombre de tabla.

VACUUM REINDEX toma considerablemente más tiempo que VACUUM FULL ya que realiza una ejecución adicional para analizar las claves de ordenación intercaladas. La operación de ordenación y fusión lleva más tiempo para tablas intercaladas porque es posible que la ordenación intercalada necesite reorganizar más filas que una ordenación compuesta.

Si una operación VACUUM REINDEX termina antes de completarse, la próxima VACUUM reanuda la operación de reindexación antes de realizar la operación de limpieza completa.

VACUUM REINDEX no puede utilizarse con TO threshold PERCENT.

RECLUSTER

Ordena las partes de la tabla que no están ordenadas. Las partes de la tabla que ya están ordenadas gracias a la ordenación automática de tabla se dejan intactas. Este comando no fusiona los datos recién ordenados con la región ordenada. Tampoco recupera todo el espacio que está marcado para su eliminación. Cuando se completa este comando, es posible que la tabla no aparezca completamente ordenada, como lo indica el campo `unsorted` en SVV_TABLE_INFO.

Recomendamos que se utilice VACUUM RECLUSTER para las tablas grandes con ingesta frecuente y consultas que solo acceden a los datos más recientes.

VACUUM RECLUSTER no es compatible con TO threshold PERCENT. Si se utiliza RECLUSTER, se necesita el nombre de una tabla.

VACUUM RECLUSTER no es compatible con las tablas con claves de ordenación intercaladas y las tablas con el estilo de distribución ALL.

table_name

El nombre de una tabla que se limpiará. Si no especifica un nombre de tabla, la operación de limpieza se aplica a todas las tablas de la base de datos actual. Puede especificar cualquier tabla permanente o temporal creada por un usuario. El comando no es significativo para otros objetos, como vistas y tablas del sistema.

Si incluye el parámetro TO threshold PERCENT, se requiere un nombre de tabla.

TO threshold PERCENT

Una cláusula que especifica el umbral por encima del cual VACUUM omite la fase de ordenación y el umbral de destino para recuperar espacio en la fase de eliminación. El umbral de ordenación es el porcentaje de las filas totales que ya están en orden en la tabla especificada antes de la limpieza.  El umbral de eliminación es el porcentaje mínimo de las filas totales que no están marcadas para eliminación después de la limpieza.

Como VACUUM reordena las filas solo cuando el porcentaje de filas ordenadas en una tabla es inferior al límite de ordenación, en muchas ocasiones, Amazon Redshift puede reducir los tiempos de VACUUM considerablemente. Del mismo modo, cuando VACUUM no se limita a recuperar el espacio del 100 por cien de las filas marcadas para su eliminación, normalmente puede omitir la reescritura de varios bloques que contienen solo algunas filas eliminadas.

Por ejemplo, si especifica 75 como umbral, VACUUM omitirá la fase de ordenación si el 75 por ciento o más de las filas de la tabla ya se encuentran en orden. En la fase de eliminación, VACUUMS establece un objetivo de espacio de recuperación de espacio en disco para que al menos el 75 por ciento de las filas de la tabla no se marquen para su eliminación después de la limpieza. El valor de threshold debe ser un número entero comprendido entre 0 y 100. El valor predeterminado es 95. Si especifica un valor de 100, VACUUM siempre ordena la tabla a menos que esté completamente ordenada y recupera espacio de todas las filas marcadas para la eliminación. Si especifica un valor de 0, VACUUM nunca ordena la tabla y nunca recupera espacio.

Si incluye el parámetro TO threshold PERCENT, también debe especificar un nombre de tabla. VACUUM falla si se omite un nombre de tabla.

No se puede utilizar el parámetro TO threshold PERCENT con REINDEX.

BOOST

Ejecuta el comando VACUUM con recursos adicionales como espacio en disco y memoria, cuando están disponibles, Con la opción BOOST, VACUUM opera en una ventana y bloquea las eliminaciones y actualizaciones simultáneas durante la duración de la operación VACUUM. La ejecución con la opción BOOST compite con los recursos del sistema, lo que puede afectar el rendimiento de las consultas. Ejecute VACUUM BOOST cuando la carga en el sistema sea ligera; por ejemplo durante las operaciones de mantenimiento.

Cuando utilice la opción BOOST, tenga en cuenta lo siguiente:

- Si se especifica BOOST, el valor table_name es obligatorio.
- BOOST no es compatible con REINDEX.
- BOOST se ignora con DELETE ONLY.

## Notas de uso

Se recomienda una limpieza completa para la mayoría de las aplicaciones de Amazon Redshift. Para obtener más información, consulte Limpieza de tablas.

Antes de ejecutar una operación de limpieza, tenga en cuenta el siguiente comportamiento:

- No se puede ejecutar VACUUM en un bloque de transacción (BEGIN … END). Para obtener más información acerca de las transacciones, consulte Aislamiento serializable.

- Parte del crecimiento de la tabla puede ocurrir cuando se limpian las tablas. Este comportamiento se espera cuando no hay filas eliminadas para recuperar o el nuevo orden de la tabla produce un menor nivel de compresión de datos.
- Durante las operaciones de limpieza, se espera cierto grado de degradación en el rendimiento de las consultas. El rendimiento normal se reanuda tan pronto se completa la operación de limpieza.
- Las operaciones de escritura simultáneas continúan durante las operaciones de limpieza, pero no se recomienda realizar operaciones de escritura durante la limpieza. Es más eficiente completar operaciones de escritura antes de ejecutar la limpieza. Además, cualquier dato que se escriba después de que se haya iniciado una operación de limpieza no se puede limpiar mediante esa operación. En este caso, es necesaria una segunda operación de limpieza.
- Es posible que una operación de limpieza no se inicie si ya hay una operación de carga o inserción en curso. Las operaciones de limpieza requieren acceso exclusivo a tablas temporalmente para poder iniciarse. Este acceso exclusivo se requiere por poco tiempo, para que las operaciones de limpieza no bloqueen cargas e inserciones simultáneas para cualquier periodo de tiempo significativo.
- Las operaciones de limpieza se omiten cuando no hay trabajo para realizar en una tabla en particular; no obstante, hay costos de mantenimiento asociados con descubrir que se puede omitir la operación. Si está seguro de que una tabla está limpia o no cumple con el umbral de limpieza, no ejecute una operación de limpieza.
- Es posible que una operación de limpieza DELETE ONLY sobre una tabla pequeña no reduzca la cantidad de bloques utilizados para almacenar los datos, especialmente cuando la tabla tiene una gran cantidad de columnas o el clúster usa una gran cantidad de sectores por nodo. Estas operaciones de limpieza agregan un bloque por columna por sector a la cuenta para inserciones simultáneas en la tabla, y hay una posibilidad de que estos costos de mantenimiento superen la reducción de bloques del espacio en disco subyacente. Por ejemplo, si una tabla de 10 columnas de un clúster de 8 nodos ocupa 1000 bloques antes de una limpieza, la limpieza no reducirá la cantidad de bloques real a menos que se recuperen más de 80 bloques de espacio en disco por la eliminación de filas. (Cada bloque de datos utiliza 1 MB).

Las operaciones de limpieza automáticas se ponen en pausa si se cumple alguna de las condiciones siguientes:

- Un usuario ejecuta una operación de lenguaje de definición de datos (DDL), como ALTER TABLE, que requiere un bloqueo exclusivo en una tabla en la que se está realizando actualmente la limpieza automática.

- Un periodo de carga elevada del clúster.

Compatibilidad con VACUUM simultáneo

Amazon Redshift admite la ejecución simultánea de múltiples transacciones de vaciado en diferentes sesiones de un clúster o grupo de trabajo. Esto significa que puede emitir diferentes y múltiples instancias de todos los modos de vaciado a la vez, con cada transacción de vaciado en una tabla única. Dos operaciones de vaciado no pueden trabajar en una misma tabla al mismo tiempo.

Directrices para ejecutar un vaciado simultáneo

- Cuando ejecute transacciones de vaciado simultáneas en diferentes sesiones, debe supervisar los recursos del sistema y evitar ejecutar demasiadas operaciones de vaciado al mismo tiempo.
- El nivel de simultaneidad recomendado depende de la cantidad de espacio que se va a recuperar, tanto del número como de la anchura de las filas que se van a ordenar, del tamaño del almacén y del tamaño de la carga de trabajo que se ejecuta junto con las operaciones VACUUM.
- Dependiendo del modo de la transacción de vaciado, comience con dos operaciones de vaciado simultáneas y agregue más dependiendo del tiempo de ejecución y carga del sistema. Al igual que otras consultas pesadas emitidas por los usuarios, las operaciones de vaciado pueden comenzar a ponerse en cola si ejecuta demasiadas simultáneamente cuando Amazon Redshift alcanza los límites de recursos del sistema.
- Ejecute varias operaciones de BOOST de vaciado con cuidado. La ejecución de vaciado con la opción BOOST compite con los recursos del sistema, lo que puede afectar el rendimiento de las consultas. Ejecute VACUUM BOOST cuando la carga en el sistema sea ligera; por ejemplo durante las operaciones de mantenimiento.
- Si no especifica un nombre de tabla, la operación de limpieza se aplica a todas las tablas de la base de datos actual. Estas operaciones de vaciado siguen ejecutándose secuencialmente.

## Ejemplos

Recuperar espacio y base de datos y reordenar las filas de todas las tablas en función del umbral de limpieza predeterminado de 95 por ciento.

```
vacuum;
```

Recuperar espacio y reordenar las filas de la tabla SALES en función del umbral predeterminado de 95 por ciento.

```
vacuum sales;
```

Siempre recuperar espacio y reordenar las filas de la tabla SALES.

```
vacuum sales to 100 percent;
```

Reordenar las filas de la tabla SALES solo si menos del 75 por ciento de las filas ya está ordenado.

```
 vacuum sort only sales to 75 percent;
```

Recuperar espacio en la tabla SALES para que al menos el 75 por ciento de las filas restantes no se marquen para su eliminación después de la limpieza.

```
vacuum delete only sales to 75 percent;
```

Reindexar y, luego, limpiar la tabla LISTING.

```
vacuum reindex listing;
```

El siguiente comando devuelve un error.

```
vacuum reindex listing to 75 percent;
```

Recluster y, a continuación, vacuum la tabla LISTING.

```
vacuum recluster listing;
```

Recluster y, a continuación, vacuum la tabla LISTING con la opción BOOST.

```
vacuum recluster listing boost;
```

# Referencia de funciones SQL

Temas


- Funciones específicas del nodo principal
- Funciones de agregación
- Funciones de matriz

- Funciones de agregación bit a bit
- Expresiones condicionales
- Funciones de formato de tipo de datos
- Funciones de fecha y hora
- Funciones hash
- Funciones HyperLogLog
- Funciones JSON
- Funciones de machine learning.
- Funciones matemáticas
- Funciones de objeto
- Funciones espaciales
- Funciones de cadena
- Funciones de información acerca del tipo SUPER
- Operadores y funciones VARBYTE
- Funciones de ventana
- Funciones de administración del sistema
- Funciones de información del sistema

Amazon Redshift admite distintas funciones que son extensiones del estándar SQL, además de funciones de agrupación estándar, funciones escalares y funciones de ventana.

> **Note**
>
> Amazon Redshift se basa en PostgreSQL. Amazon Redshift y PostgreSQL tienen una serie de diferencias muy importantes que debe tener en cuenta al diseñar y desarrollar aplicaciones de almacenamiento de datos. Para obtener más información acerca de las diferencias entre Amazon Redshift SQL y PostgreSQL, consulte Amazon Redshift y PostgreSQL.

## Funciones específicas del nodo principal

Algunas consultas de Amazon Redshift se distribuyen y ejecutan en los nodos de computación, en tanto que otras lo hacen exclusivamente en el nodo principal.

El nodo principal distribuye los datos a los nodos de computación cuando una consulta utiliza como referencia tablas o tablas de sistemas creadas por el usuario (tablas con un prefijo STL o STV y vistas de sistema con un prefijo SVL o SVV). Una consulta que utiliza como referencia solamente tablas de catálogos (tablas con un prefijo PG, como por ejemplo PG_TABLE_DEF) o que no utiliza ninguna tabla se ejecuta únicamente en el nodo principal.

Algunas funciones SQL de Amazon Redshift solo se admiten en el nodo principal, y no en los nodos informáticos. Una consulta que utiliza una función del nodo principal debe ejecutarse exclusivamente en ese nodo, no en los nodos informáticos, ya que se produciría un error.

En la documentación de cada función específica del nodo principal, se incluye una nota en la que se indica que la función producirá un error si utiliza como referencia tablas definidas por el usuario o tablas del sistema de Amazon Redshift.

Para obtener más información, consulte Funciones SQL admitidas en el nodo principal.

Las siguientes funciones de SQL son específicas del nodo principal y no son compatibles con los nodos de computación:

Funciones de información del sistema

- CURRENT_SCHEMA
- CURRENT_SCHEMAS
- HAS_DATABASE_PRIVILEGE
- HAS_SCHEMA_PRIVILEGE
- HAS_TABLE_PRIVILEGE

Funciones de cadena

- SUBSTR

Funciones matemáticas

- FACTORIAL()

Las siguientes funciones específicas del nodo principal están obsoletas y ya no son compatibles:

Funciones de datos

- AGE
- CURRENT_TIME
- CURRENT_TIMESTAMP
- LOCALTIME
- ISFINITE
- NOW

Funciones de cadena

- GETBIT
- GET_BYTE
- SET_BIT
- SET_BYTE
- TO_ASCII

## Funciones de agregación

Temas


- Función ANY_VALUE
- Función APPROXIMATE PERCENTILE_DISC
- Función AVG
- Función COUNT
- Función LISTAGG
- Función MAX
- Función MEDIAN
- Función MIN
- Función PERCENTILE_CONT
- Funciones STDDEV_SAMP y STDDEV_POP
- Función SUM
- Funciones VAR_SAMP y VAR_POP

Las funciones de agregación computan un valor de resultado único correspondiente a un conjunto de valores de entrada.

Las instrucciones SELECT que utilizan funciones de agregación pueden incluir dos cláusulas opcionales: GROUP BY y HAVING. La sintaxis para estas cláusulas es la siguiente (utilizando la función COUNT como ejemplo):

```
SELECT count (*) expression FROM table_reference
WHERE condition [GROUP BY expression ] [ HAVING condition]
```

La cláusula GROUP BY agrega y agrupa resultados por valores únicos en columnas específicas. La cláusula HAVING restringe los resultados de retorno en filas donde una condición de agregación en particular es true, como por ejemplo count (*) > 1. La cláusula HAVING se usa de la misma manera que WHERE: para restringir filas según el valor de una columna. Consulte COUNT para ver un ejemplo de estas cláusulas adicionales.

Las funciones de agregación no aceptan como argumentos funciones de agregación anidadas ni funciones de ventana.

## Función ANY_VALUE

La función ANY_VALUE devuelve cualquier valor de los valores de expresión de entrada de una manera que no sea determinista. Esta función devuelve NULL si la expresión de entrada no implica que se devuelva alguna fila. La función también puede devolver NULL si hay valores NULL en la expresión de entrada.

Sintaxis

```
ANY_VALUE( [ DISTINCT | ALL ] expression )
```

Argumentos

DISTINCT | ALL

Especifique DISTINCT u ALL para devolver cualquier valor de los valores de expresión de entrada. El argumento DISTINCT no tiene ningún efecto y se pasa por alto.

expression

La columna o la expresión de destino en la que opera la función. La expresión corresponde a uno de los siguientes tipos de datos:

- SMALLINT
- INTEGER
- BIGINT
- DECIMAL
- REAL
- DOUBLE PRECISION
- BOOLEAN
- CHAR
- VARCHAR
- DATE
- TIMESTAMP
- TIMESTAMPTZ
- TIME
- TIMETZ
- INTERVALO AÑO A MES
- INTERVALO DÍA A SEGUNDO
- VARBYTE
- SUPER
- HLLSKETCH
- GEOMETRY
- GEOGRAPHY

Devuelve

Devuelve el mismo tipo de datos que expresión.

Notas de uso

Si una instrucción que especifica la función ANY_VALUE para una columna también incluye una segunda referencia de columna, la segunda columna debe aparecer en una cláusula GROUP BY o debe incluirse en una función de agrupación.

### Ejemplos

En los ejemplos, se utiliza la tabla de eventos que se crea en el Paso 4: Cargar datos de muestra desde Amazon S3 en la Guía de introducción a Amazon Redshift. El siguiente ejemplo devuelve una instancia de cualquier DATEID donde el nombre del evento es Eagles.

```
select any_value(dateid) as dateid, eventname from event where eventname ='Eagles'
 group by eventname;
```

A continuación, se muestran los resultados.

```
dateid | eventname
-------+---------------
 1878  | Eagles
```

El siguiente ejemplo devuelve una instancia de cualquier DATEID donde el nombre del evento es Eagles o Cold War Kids.

```
select any_value(dateid) as dateid, eventname from event where eventname in('Eagles',
 'Cold War Kids') group by eventname;
```

A continuación, se muestran los resultados.

```
dateid | eventname
-------+---------------
 1922  | Cold War Kids
 1878  | Eagles
```

## Función APPROXIMATE PERCENTILE_DISC

APPROXIMATE PERCENTILE_DISC es una función de distribución inversa que asume un modelo de distribución discreta. Toma un valor percentil y una especificación de ordenación, y devuelve un elemento del conjunto dado. La aproximación permite que la función se ejecute mucho más rápido, con un margen de error relativamente bajo de alrededor del 0,5 %.

En un percentil dado, APPROXIMATE PERCENTILE_DISC utiliza un algoritmo de resumen de cuartiles para aproximar el percentil discreto de la expresión en la cláusula ORDER BY. APPROXIMATE PERCENTILE_DISC devuelve el menor valor de distribución acumulado (con respecto a la misma especificación de ordenación) que sea mayor o igual que el percentil.

### Sintaxis

```
APPROXIMATE  PERCENTILE_DISC ( percentile )
WITHIN GROUP (ORDER BY expr)
```

### Argumentos

percentil

Constante numérica entre 0 y 1. Los valores nulos se ignoran en el cálculo.

WITHIN GROUP ( ORDER BY expr)

Cláusula que especifica valores numéricos o de fecha/hora para ordenar y calcular el percentil.

### Devuelve

El mismo tipo de datos que la expresión ORDER BY en la cláusula WITHIN GROUP.

### Notas de uso

Si la instrucción APPROXIMATE PERCENTILE_DISC incluye una cláusula GROUP BY, el conjunto de resultados es limitado. El límite varía según el tipo de nodo y la cantidad de nodos. Si se supera el límite, la función falla y devuelve el siguiente mensaje de error.

```
GROUP BY limit for approximate percentile_disc exceeded.
```

Si necesita evaluar más grupos que los permitidos, considere usar Función PERCENTILE_CONT.

### Ejemplos

El siguiente ejemplo devuelve la cantidad de ventas, las ventas totales y el valor del décimo quinto percentil para las 10 primeras fechas.

```
select top 10 date.caldate,
count(totalprice), sum(totalprice),
approximate percentile_disc(0.5)
within group (order by totalprice)
from listing
join date on listing.dateid = date.dateid
group by date.caldate
```

```
order by 3 desc;

caldate    | count | sum        | percentile_disc
-----------+-------+------------+----------------
2008-01-07 |   658 | 2081400.00 |         2020.00
2008-01-02 |   614 | 2064840.00 |         2178.00
2008-07-22 |   593 | 1994256.00 |         2214.00
2008-01-26 |   595 | 1993188.00 |         2272.00
2008-02-24 |   655 | 1975345.00 |         2070.00
2008-02-04 |   616 | 1972491.00 |         1995.00
2008-02-14 |   628 | 1971759.00 |         2184.00
2008-09-01 |   600 | 1944976.00 |         2100.00
2008-07-29 |   597 | 1944488.00 |         2106.00
2008-07-23 |   592 | 1943265.00 |         1974.00
```

## Función AVG

La función AVG devuelve el promedio (media aritmética) de los valores de la expresión de entrada. La función AVG funciona con valores numéricos e ignora los valores NULL.

### Sintaxis

```
AVG ( [ DISTINCT | ALL ] expression )
```

### Argumentos

expression

La columna o expresión de destino sobre la que opera la función. La expresión corresponde a uno de los siguientes tipos de datos:

- SMALLINT
- INTEGER
- BIGINT
- NUMERIC
- DECIMAL
- REAL
- DOUBLE PRECISION
- SUPER

DISTINCT | ALL

Con el argumento DISTINCT, la función elimina todos los valores duplicados de la expresión especificada antes de calcular el promedio. Con el argumento ALL, la función retiene todos los valores duplicados de la expresión especificada para calcular el promedio. El valor predeterminado es ALL.

Tipos de datos

Los tipos de argumento compatibles con la función AVG son SMALLINT, INTEGER, BIGINT, NUMERIC, DECIMAL, REAL, DOUBLE PRECISION y SUPER.

Los tipos de retorno compatibles con la función AVG son los siguientes:

- BIGINT para cualquier argumento de tipo entero
- DOUBLE PRECISION para un argumento de punto flotante
- Devuelve el mismo tipo de datos como expresión para cualquier otro tipo de argumento.

La precisión predeterminada para un resultado de la función AVG con un argumento NUMERIC o DECIMAL de es 38. La escala del resultado es la misma que la escala del argumento. Por ejemplo, una función AVG de una columna DEC(5,2) devuelve un tipo de datos DEC(38,2).

Ejemplos

Encontrar la cantidad promedio vendida por transacción de la tabla SALES:

```
select avg(qtysold)from sales;

avg
-----
2
(1 row)
```

Encontrar el precio total promedio en lista para todas las listas:

```
select avg(numtickets*priceperticket) as avg_total_price from listing;

avg_total_price
-----------------
```

```
3034.41
(1 row)
```

Encontrar el precio promedio pagado, agrupado por mes en orden descendiente:

```
select avg(pricepaid) as avg_price, month
from sales, date
where sales.dateid = date.dateid
group by month
order by avg_price desc;

avg_price | month
-----------+-------
659.34 | MAR
655.06 | APR
645.82 | JAN
643.10 | MAY
642.72 | JUN
642.37 | SEP
640.72 | OCT
640.57 | DEC
635.34 | JUL
635.24 | FEB
634.24 | NOV
632.78 | AUG
(12 rows)
```

## Función COUNT

La función COUNT cuenta las filas definidas por la expresión.

La función COUNT tiene las siguientes variaciones.

- COUNT ( * ) cuenta todas las filas en la tabla destino, incluya o no valores nulos.
- COUNT ( expresión ) calcula el número de filas con valores no NULL de una determinada columna o expresión.
- COUNT ( DISTINCT expresión ) calcula el número de valores no NULL diferentes de una columna o expresión.
- APPROXIMATE COUNT DISTINCT realiza un cálculo aproximado del número de valores distintos de NULL diferentes de una columna o una expresión.

### Sintaxis

```
COUNT( * | expression )
```

```
COUNT ( [ DISTINCT | ALL ] expression )
```

```
APPROXIMATE COUNT ( DISTINCT expression )
```

### Argumentos

expression

La columna o expresión de destino sobre la que opera la función. La función COUNT admite todos los tipos de datos de argumentos.

DISTINCT | ALL

Con el argumento DISTINCT, la función elimina todos los valores duplicados de la expresión especificada antes de hacer el conteo. Con el argumento ALL, la función retiene todos los valores duplicados de la expresión para el conteo. El valor predeterminado es ALL.

APPROXIMATE

Cuando se usa con APPROXIMATE, la función COUNT DISTINCT utiliza un algoritmo HyperLogLog para realizar un cálculo aproximado del número de valores distintos de NULL de una columna o una expresión. Las consultas que utilizan la palabra clave APPROXIMATE se ejecutan mucho más rápido, con un margen de error relativamente bajo de alrededor del 2 %. La aproximación se justifica para consultas que devuelven una cantidad grande de valores distintos, millones o más por consulta, o por grupo, si es que hay una cláusula de grupo. Para conjuntos más pequeños de valores distintos, en miles, la aproximación puede ser más lenta que un conteo preciso. APPROXIMATE solo se puede usar con COUNT DISTINCT.

### Tipo de retorno

La función COUNT devuelve BIGINT.

### Ejemplos

Cuente todos los usuarios del estado de Florida:

```
select count(*) from users where state='FL';
```

```
count
-------
510
```

Cuenta todos los nombres de evento de la tabla EVENT:

```
select count(eventname) from event;

count
-------
8798
```

Cuenta todos los nombres de evento de la tabla EVENT:

```
select count(all eventname) from event;

count
-------
8798
```

Contar todos los ID únicos de lugares de la tabla EVENT:

```
select count(distinct venueid) as venues from event;

venues
--------
204
```

Contar la cantidad de veces que cada vendedor indicó lotes de más de cuatro tickets para venta. Agrupar los resultados según ID de vendedor:

```
select count(*), sellerid from listing
where numtickets > 4
group by sellerid
order by 1 desc, 2;

count | sellerid
------+----------
12    |    6386
11    |    17304
```

```
11    |    20123
11    |    25428
...
```

En los siguientes ejemplos, se comparan los valores de retorno y los tiempos de ejecución para COUNT y APPROXIMATE COUNT.

```
select  count(distinct pricepaid) from sales;

count
-------
  4528


Time: 48.048 ms


select approximate count(distinct pricepaid) from sales;

count
-------
  4553


Time: 21.728 ms
```

## Función LISTAGG

Para cada grupo de una consulta, la función de agregación LISTAGG ordena las filas de ese grupo según la expresión ORDER BY y, luego, concatena los valores en una sola cadena.

### Sintaxis

```
LISTAGG( [DISTINCT] aggregate_expression [, 'delimiter' ] )
[ WITHIN GROUP (ORDER BY order_list) ]
```

### Argumentos

#### DISTINCT

Una cláusula que elimina los valores duplicados de la expresión especificada antes de la concatenación. Se omiten los espacios finales. Por ejemplo, las cadenas `'a'` y `'a '` se tratan

como duplicados. LISTAGG usa el primer valor que se encuentra. Para obtener más información, consulte Importancia de los espacios en blancos anteriores y posteriores.

expresión_de_agregación

Toda expresión válida, como un nombre de columna, que proporcione los valores para la agregación. Se ignoran los valores NULL y las cadenas vacías.

delimiter

La constante de cadena que separa los valores concatenados. El valor predeterminado es NULL.

WITHIN GROUP (ORDER BY order_list)

Una cláusula que especifica el orden de los valores agregados.

Devuelve

VARCHAR(MAX). Si el conjunto de resultados es mayor que el tamaño máximo VARCHAR, LISTAGG devuelve el siguiente error:

```
Invalid operation: Result size exceeds LISTAGG limit
```

Notas de uso

- Si una instrucción incluye varias funciones LISTAGG que usan cláusulas WITHIN GROUP, cada cláusula WITHIN GROUP debe usar los mismos valores ORDER BY.

  Por ejemplo, la siguiente instrucción devuelve un error.

  ```
  SELECT LISTAGG(sellerid)
  WITHIN GROUP (ORDER BY dateid) AS sellers,
  LISTAGG(dateid)
  WITHIN GROUP (ORDER BY sellerid) AS dates
  FROM sales;
  ```

  Las siguientes instrucciones se ejecutan correctamente.

  ```
  SELECT LISTAGG(sellerid)
  WITHIN GROUP (ORDER BY dateid) AS sellers,
  LISTAGG(dateid)
  WITHIN GROUP (ORDER BY dateid) AS dates
  FROM sales;
  ```

```
SELECT LISTAGG(sellerid)
WITHIN GROUP (ORDER BY dateid) AS sellers,
LISTAGG(dateid) AS dates
FROM sales;
```

- No puede utilizar las funciones agregadas LISTAGG, PERCENTILE_CONT y MEDIAN con otras funciones agregadas distintas.

### Ejemplos

En el siguiente ejemplo, se agrega el ID de vendedores ordenados por ID de vendedor.

```
SELECT LISTAGG(sellerid, ', ')
WITHIN GROUP (ORDER BY sellerid)
FROM sales
WHERE eventid = 4337;

listagg

----------------------------------------------------------------------------------------------------
380, 380, 1178, 1178, 1178, 2731, 8117, 12905, 32043, 32043, 32043, 32432, 32432,
 38669, 38750, 41498, 45676, 46324, 47188, 47188, 48294
```

En el siguiente ejemplo, se usa DISTINCT para devolver una lista de ID de vendedor distintos.

```
SELECT LISTAGG(DISTINCT sellerid, ', ')
WITHIN GROUP (ORDER BY sellerid)
FROM sales
WHERE eventid = 4337;

listagg

------------------------------------------------------------------------------------------
380, 1178, 2731, 8117, 12905, 32043, 32432, 38669, 38750, 41498, 45676, 46324, 47188,
 48294
```

En el siguiente ejemplo, se agrega el ID de vendedores ordenados por fecha.

```
SELECT LISTAGG(sellerid, ', ')
WITHIN GROUP (ORDER BY dateid)
FROM sales
```

```
WHERE eventid = 4337;

   listagg
-----------------------------------------------------------------------------------------------
 41498, 47188, 47188, 1178, 1178, 1178, 380, 45676, 46324, 48294, 32043, 32043, 32432,
 12905, 8117, 38750, 2731, 32432, 32043, 380, 38669
```

El siguiente ejemplo devuelve una lista separada con barras de fechas de ventas para el comprador con un ID de 660.

```
SELECT LISTAGG(
    (SELECT caldate FROM date WHERE date.dateid=sales.dateid), ' | '
)
WITHIN GROUP (ORDER BY sellerid DESC, salesid ASC)
FROM sales
WHERE buyerid = 660;

             listagg
-------------------------------------------------
2008-07-16 | 2008-07-09 | 2008-01-01 | 2008-10-26
```

El siguiente ejemplo devuelve una lista separada por comas de ID de ventas para los ID de comprador 660, 661 y 662.

```
SELECT buyerid,
LISTAGG(salesid,', ')
WITHIN GROUP (ORDER BY salesid) AS sales_id
FROM sales
WHERE buyerid BETWEEN 660 AND 662
GROUP BY buyerid
ORDER BY buyerid;

buyerid |                sales_id
--------+-------------------------------------------------
660     | 32872, 33095, 33514, 34548
661     | 19951, 20517, 21695, 21931
662     | 3318, 3823, 4215, 51980, 53202, 55908, 57832, 171603
```

## Función MAX

La función MAX devuelve el valor máximo en un conjunto de filas. Es posible utilizar DISTINCT o ALL pero no influye en el resultado.

## Sintaxis

```
MAX ( [ DISTINCT | ALL ] expression )
```

## Argumentos

### expression

La columna o expresión de destino sobre la que opera la función. La expresión corresponde a uno de los siguientes tipos de datos:

- SMALLINT
- INTEGER
- BIGINT
- DECIMAL
- REAL
- DOUBLE PRECISION
- CHAR
- VARCHAR
- DATE
- TIMESTAMP
- TIMESTAMPTZ
- TIME
- TIMETZ
- VARBYTE
- SUPER

### DISTINCT | ALL

Con el argumento DISTINCT, la función elimina todos los valores duplicados de la expresión especificada antes de calcular el máximo. Con el argumento ALL, la función retiene todos los valores duplicados de la expresión especificada para calcular el máximo. El valor predeterminado es ALL.

Tipos de datos

Devuelve el mismo tipo de datos que expresión. El equivalente en valor booleano de la función MIN es Función BOOL_AND y el equivalente en booleano de MAX es Función BOOL_OR.

Ejemplos

Encontrar el precio más alto pagado de todas las ventas:

```
select max(pricepaid) from sales;

max
----------
12624.00
(1 row)
```

Encontrar el precio más alto pagado por ticket de todas las ventas:

```
select max(pricepaid/qtysold) as max_ticket_price
from sales;

max_ticket_price
-----------------
2500.00000000
(1 row)
```

## Función MEDIAN

Calcula el valor medio del rango de valores. Los valores NULL del rango se ignoran.

MEDIAN es una función de distribución inversa que asume un modelo de distribución continua.

MEDIAN es un caso especial de PERCENTILE_CONT.

Sintaxis

```
MEDIAN(median_expression)
```

Argumentos

expresión_de_mediana

La columna o expresión de destino sobre la que opera la función.

Tipos de datos

El tipo de valor devuelto viene determinado por el tipo de datos de expresión_de_mediana. En la tabla siguiente, se muestra el tipo de valor devuelto para cada tipo de datos de expresión_de_mediana.

| Tipo de entrada | Tipo de retorno |
| --- | --- |
| INT2, INT4, INT8, NUMERIC, DECIMAL | DECIMAL |
| FLOAT, DOUBLE | DOUBLE |
| DATE | DATE |
| TIMESTAMP | TIMESTAMP |
| TIMESTAMPTZ | TIMESTAMPTZ |

Notas de uso

Si el argumento median_expression es un tipo de dato DECIMAL definido con la precisión máxima de 38 dígitos, es posible que MEDIAN devuelva un resultado impreciso o un error. Si el valor de retorno de la función MEDIAN supera los 38 dígitos, el resultado se trunca para adaptarse, lo que genera pérdida de precisión. Si, durante la interpolación, un resultado intermedio supera la precisión máxima, se produce un desbordamiento numérico y la función devuelve un error. Para evitar estas condiciones, recomendamos usar un tipo de dato con menor precisión o emitir el argumento median_expression con una precisión menor.

Si una instrucción incluye varias llamadas a funciones de agregación basadas en ordenación (LISTAGG, PERCENTILE_CONT o MEDIAN), todas deben usar los mismos valores ORDER BY. Tenga en cuenta que MEDIAN aplica un comando ORDER BY implícito en el valor de expresión.

Por ejemplo, la siguiente instrucción devuelve un error.

```
SELECT TOP 10 salesid, SUM(pricepaid),
PERCENTILE_CONT(0.6) WITHIN GROUP(ORDER BY salesid),
MEDIAN(pricepaid)
FROM sales
GROUP BY salesid, pricepaid;
```

```
An error occurred when executing the SQL command:
SELECT TOP 10 salesid, SUM(pricepaid),
PERCENTILE_CONT(0.6) WITHIN GROUP(ORDER BY salesid),
MEDIAN(pricepaid)
FROM sales
GROUP BY salesid, pricepaid;

ERROR: within group ORDER BY clauses for aggregate functions must be the same
```

La siguiente instrucción se ejecuta correctamente.

```
SELECT TOP 10 salesid, SUM(pricepaid),
PERCENTILE_CONT(0.6) WITHIN GROUP(ORDER BY salesid),
MEDIAN(salesid)
FROM sales
GROUP BY salesid, pricepaid;
```

Ejemplos

En los siguientes ejemplos se utiliza la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

En el siguiente ejemplo, se muestra que MEDIAN produce los mismos resultados que PERCENTILE_CONT(0.5).

```
SELECT TOP 10 DISTINCT sellerid, qtysold,
PERCENTILE_CONT(0.5) WITHIN GROUP(ORDER BY qtysold),
MEDIAN(qtysold)
FROM sales
GROUP BY sellerid, qtysold;

+----------+---------+-----------------+--------+
| sellerid | qtysold | percentile_cont | median |
+----------+---------+-----------------+--------+
|        2 |       2 |               2 |      2 |
|       26 |       1 |               1 |      1 |
|       33 |       1 |               1 |      1 |
|       38 |       1 |               1 |      1 |
|       43 |       1 |               1 |      1 |
|       48 |       2 |               2 |      2 |
|       48 |       3 |               3 |      3 |
```

```
|       77 |       4 |                4 |      4 |
|       85 |       4 |                4 |      4 |
|       95 |       2 |                2 |      2 |
+----------+---------+------------------+--------+
```

En el siguiente ejemplo se encuentra la cantidad media vendida para cada sellerid.

```
SELECT sellerid,
MEDIAN(qtysold)
FROM sales
GROUP BY sellerid
ORDER BY sellerid
LIMIT 10;

+----------+--------+
| sellerid | median |
+----------+--------+
|        1 |    1.5 |
|        2 |      2 |
|        3 |      2 |
|        4 |      2 |
|        5 |      1 |
|        6 |      1 |
|        7 |    1.5 |
|        8 |      1 |
|        9 |      4 |
|       12 |      2 |
+----------+--------+
```

Para comprobar los resultados de la consulta anterior del primer sellerid, use el siguiente ejemplo.

```
SELECT qtysold
FROM sales
WHERE sellerid=1;

+---------+
| qtysold |
+---------+
|       2 |
|       1 |
+---------+
```

## Función MIN

La función MIN devuelve el valor mínimo en un conjunto de filas. Es posible utilizar DISTINCT o ALL pero no influye en el resultado.

### Sintaxis

```
MIN ( [ DISTINCT | ALL ] expression )
```

### Argumentos

expression

La columna o expresión de destino sobre la que opera la función. La expresión corresponde a uno de los siguientes tipos de datos:

- SMALLINT
- INTEGER
- BIGINT
- DECIMAL
- REAL
- DOUBLE PRECISION
- CHAR
- VARCHAR
- DATE
- TIMESTAMP
- TIMESTAMPTZ
- TIME
- TIMETZ
- VARBYTE
- SUPER

DISTINCT | ALL

Con el argumento DISTINCT, la función elimina todos los valores duplicados de la expresión especificada antes de calcular el mínimo. Con el argumento ALL, la función retiene todos los

valores duplicados de la expresión especificada para calcular el mínimo. El valor predeterminado es ALL.

Tipos de datos

Devuelve el mismo tipo de datos que expresión. El equivalente en valor booleano de la función MIN es Función BOOL_AND y el equivalente en booleano de MAX es Función BOOL_OR.

Ejemplos

Encontrar el precio más bajo pagado de todas las ventas:

```
select min(pricepaid) from sales;

min
-------
20.00
(1 row)
```

Encontrar el precio más bajo pagado por ticket de todas las ventas:

```
select min(pricepaid/qtysold)as min_ticket_price
from sales;

min_ticket_price
------------------
20.00000000
(1 row)
```

## Función PERCENTILE_CONT

PERCENTILE_CONT es una función de distribución inversa que asume un modelo de distribución continua. Toma un valor percentil y una especificación de ordenación, y devuelve un valor interpolado que se encuadraría dentro de ese valor percentil dado respecto de la especificación de ordenación.

PERCENTILE_CONT computa una interpolación linear entre valores después de ordenarlos. Utilizando el valor percentil `(P)` y la cantidad de filas no nulas `(N)` en el grupo de agregación, la función computa la cantidad de filas después de ordenar las filas según la especificación de ordenación. Esta cantidad de filas `(RN)` se computa según la fórmula `RN = (1+ (P*(N-1))`. El

resultado final de la función de agregación se computa por interpolación lineal entre los valores de filas en números de filas `CRN = CEILING(RN)` y `FRN = FLOOR(RN)`.

El resultado final será el siguiente.

Si `(CRN = FRN = RN)`, entonces el resultado es `(value of expression from row at RN)`

De lo contrario, el resultado es el siguiente:

`(CRN - RN) * (value of expression for row at FRN) + (RN - FRN) * (value of expression for row at CRN)`.

Sintaxis

```
PERCENTILE_CONT(percentile)
WITHIN GROUP(ORDER BY expr)
```

Argumentos

percentil

Constante numérica entre 0 y 1. Los valores NULL se ignoran en el cálculo.

expr

Valores numéricos o de fecha/hora específicos para ordenar y calcular el percentil.

Devuelve

El tipo de retorno se determina por el tipo de datos de la expresión ORDER BY en la cláusula WITHIN GROUP. En la tabla siguiente, se muestra el tipo de retorno para cada tipo de datos de la expresión ORDER BY.

| Tipo de entrada | Tipo de retorno |
|---|---|
| INT2, INT4, INT8, NUMERIC, DECIMAL | DECIMAL |
| FLOAT, DOUBLE | DOUBLE |
| DATE | DATE |

| Tipo de entrada | Tipo de retorno |
| --- | --- |
| TIMESTAMP | TIMESTAMP |
| TIMESTAMPTZ | TIMESTAMPTZ |

Notas de uso

Si la expresión ORDER BY es un tipo de dato DECIMAL definido con la precisión máxima de 38 dígitos, es posible que PERCENTILE_CONT devuelva un resultado impreciso o un error. Si el valor de retorno de la función PERCENTILE_CONT supera los 38 dígitos, el resultado se trunca para adaptarse, lo que genera pérdida de precisión. Si, durante la interpolación, un resultado intermedio supera la precisión máxima, se produce un desbordamiento numérico y la función devuelve un error. Para evitar estas condiciones, recomendamos usar un tipo de dato con menor precisión o emitir la expresión ORDER BY a una precisión menor.

Si una instrucción incluye varias llamadas a funciones de agregación basadas en ordenación (LISTAGG, PERCENTILE_CONT o MEDIAN), todas deben usar los mismos valores ORDER BY. Tenga en cuenta que MEDIAN aplica un comando ORDER BY implícito en el valor de expresión.

Por ejemplo, la siguiente instrucción devuelve un error.

```
SELECT TOP 10 salesid, SUM(pricepaid),
PERCENTILE_CONT(0.6) WITHIN GROUP(ORDER BY salesid),
MEDIAN(pricepaid)
FROM sales
GROUP BY salesid, pricepaid;

An error occurred when executing the SQL command:
SELECT TOP 10 salesid, SUM(pricepaid),
PERCENTILE_CONT(0.6) WITHIN GROUP(ORDER BY salesid),
MEDIAN(pricepaid)
FROM sales
GROUP BY salesid, pricepaid;

ERROR: within group ORDER BY clauses for aggregate functions must be the same
```

La siguiente instrucción se ejecuta correctamente.

```
SELECT TOP 10 salesid, SUM(pricepaid),
```

```
PERCENTILE_CONT(0.6) WITHIN GROUP(ORDER BY salesid),
MEDIAN(salesid)
FROM sales
GROUP BY salesid, pricepaid;
```

Ejemplos

En los siguientes ejemplos se utiliza la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

En el siguiente ejemplo, se muestra que PERCENTILE_CONT(0.5) produce los mismos resultados que MEDIAN.

```
SELECT TOP 10 DISTINCT sellerid, qtysold,
PERCENTILE_CONT(0.5) WITHIN GROUP(ORDER BY qtysold),
MEDIAN(qtysold)
FROM sales
GROUP BY sellerid, qtysold;

+----------+---------+-----------------+--------+
| sellerid | qtysold | percentile_cont | median |
+----------+---------+-----------------+--------+
|        2 |       2 |               2 |      2 |
|       26 |       1 |               1 |      1 |
|       33 |       1 |               1 |      1 |
|       38 |       1 |               1 |      1 |
|       43 |       1 |               1 |      1 |
|       48 |       2 |               2 |      2 |
|       48 |       3 |               3 |      3 |
|       77 |       4 |               4 |      4 |
|       85 |       4 |               4 |      4 |
|       95 |       2 |               2 |      2 |
+----------+---------+-----------------+--------+
```

En el siguiente ejemplo, se encuentra PERCENTILE_CONT(0.5) y PERCENTILE_CONT(0.75) para la cantidad vendida por cada sellerid en la tabla SALES.

```
SELECT sellerid,
PERCENTILE_CONT(0.5) WITHIN GROUP(ORDER BY qtysold) as pct_50,
PERCENTILE_CONT(0.75) WITHIN GROUP(ORDER BY qtysold) as pct_75
FROM sales
GROUP BY sellerid
ORDER BY sellerid
```

```
LIMIT 10;

+----------+--------+---------+
| sellerid | pct_50 | pct_75 |
+----------+--------+---------+
|        1 |    1.5 |    1.75 |
|        2 |      2 |    2.25 |
|        3 |      2 |       3 |
|        4 |      2 |       2 |
|        5 |      1 |     1.5 |
|        6 |      1 |       1 |
|        7 |    1.5 |    1.75 |
|        8 |      1 |       1 |
|        9 |      4 |       4 |
|       12 |      2 |    3.25 |
+----------+--------+---------+
```

Para comprobar los resultados de la consulta anterior del primer sellerid, use el siguiente ejemplo.

```
SELECT qtysold
FROM sales
WHERE sellerid=1;

+---------+
| qtysold |
+---------+
|       2 |
|       1 |
+---------+
```

## Funciones STDDEV_SAMP y STDDEV_POP

Las funciones STDDEV_SAMP y STDDEV_POP devuelven la muestra y la desviación estándar de población de un conjunto de valores numéricos (entero, decimal o de punto flotante). El resultado de la función STDDEV_SAMP es equivalente a la raíz cuadrada de la varianza de muestra del mismo conjunto de valores.

STDDEV_SAMP y STDDEV son sinónimos para la misma función.

### Sintaxis

```
STDDEV_SAMP | STDDEV ( [ DISTINCT | ALL ] expression)
```

```
STDDEV_POP ( [ DISTINCT | ALL ] expression)
```

La expresión debe ser un tipo de datos entero, decimal o de punto flotante. Independientemente del tipo de datos de la expresión, el tipo de retorno de esta función es un número de doble precisión.

> **Note**
>
> La desviación estándar se calcula utilizando aritmética de punto flotante, que puede dar como resultado una leve imprecisión.

Notas de uso

Cuando la desviación estándar de la muestra (STDDEV o STDDEV_SAMP) se calcula para una expresión que consta de un valor único, el resultado de la función es NULL no 0.

Ejemplos

La siguiente consulta devuelve el promedio de valores en la columna VENUESEATS de la tabla VENUE, seguido de la desviación estándar de la muestra y la desviación estándar de la población del mismo conjunto de valores. VENUESEATS es una columna INTEGER. La escala del resultado se reduce a 2 dígitos.

```
select avg(venueseats),
cast(stddev_samp(venueseats) as dec(14,2)) stddevsamp,
cast(stddev_pop(venueseats) as dec(14,2)) stddevpop
from venue;

avg  | stddevsamp | stddevpop
-------+------------+-----------
17503 |   27847.76 |  27773.20
(1 row)
```

La siguiente consulta devuelve la desviación estándar de muestra para la columna COMMISSION en la tabla SALES. COMMISSION es una columna DECIMAL. La escala del resultado se reduce a 10 dígitos.

```
select cast(stddev(commission) as dec(18,10))
from sales;
```

```
stddev
----------------
130.3912659086
(1 row)
```

La siguiente consulta convierte la desviación estándar de muestra para la columna COMMISSION en un número entero.

```
select cast(stddev(commission) as integer)
from sales;

stddev
--------
130
(1 row)
```

La siguiente consulta devuelve tanto la desviación estándar de muestra y la raíz cuadrada de la varianza de muestra para la columna COMMISSION. Los resultados de estos cálculos son semejantes.

```
select
cast(stddev_samp(commission) as dec(18,10)) stddevsamp,
cast(sqrt(var_samp(commission)) as dec(18,10)) sqrtvarsamp
from sales;

stddevsamp   |  sqrtvarsamp
----------------+----------------
130.3912659086 | 130.3912659086
(1 row)
```

## Función SUM

La función SUM devuelve la suma de la columna de entrada o valores de la expresión. La función SUM funciona con valores numéricos e ignora los valores NULL.

Sintaxis

```
SUM ( [ DISTINCT | ALL ] expression )
```

### Argumentos

expression

La columna o expresión de destino sobre la que opera la función. La expresión corresponde a uno de los siguientes tipos de datos:

- SMALLINT
- INTEGER
- BIGINT
- NUMERIC
- DECIMAL
- REAL
- DOUBLE PRECISION
- SUPER

DISTINCT | ALL

Con el argumento DISTINCT, la función elimina todos los valores duplicados de la expresión especificada antes de calcular la suma. Con el argumento ALL, la función retiene todos los valores duplicados de la expresión especificada para calcular la suma. El valor predeterminado es ALL.

### Tipos de datos

Los tipos de argumento compatibles con la función SUM son SMALLINT, INTEGER, BIGINT, NUMERIC, DECIMAL, REAL, DOUBLE PRECISION y SUPER.

Los tipos de retorno compatibles con la función SUM son los siguientes:

- BIGINT para los argumentos BIGINT, SMALLINT e INTEGER
- NUMERIC para argumentos NUMERIC
- DOUBLE PRECISION para argumentos de punto flotante
- Devuelve el mismo tipo de datos como expresión para cualquier otro tipo de argumento.

La precisión predeterminada para un resultado de la función SUM con un argumento NUMERIC o DECIMAL es 38. La escala del resultado es la misma que la escala del argumento. Por ejemplo, un comando SUM de una columna DEC(5,2) devuelve un tipo de dato DEC(38,2).

### Ejemplos

Encontrar la suma de todas las comisiones pagadas de la tabla SALES:

```
select sum(commission) from sales;

sum
-------------
16614814.65
(1 row)
```

Encontrar la cantidad de asientos en todos los lugares en el estado de Florida:

```
select sum(venueseats) from venue
where venuestate = 'FL';

sum
--------
250411
(1 row)
```

Encontrar la cantidad de asientos vendidos en mayo:

```
select sum(qtysold) from sales, date
where sales.dateid = date.dateid and date.month = 'MAY';

sum
-------
32291
(1 row)
```

## Funciones VAR_SAMP y VAR_POP

Las funciones VAR_SAMP y VAR_POP devuelven la muestra y la varianza de población de un conjunto de valores numéricos (entero, decimal o de punto flotante). El resultado de la función VAR_SAMP es equivalente a la desviación cuadrada estándar de la muestra del mismo conjunto de valores.

VAR_SAMP y VARIANCE son sinónimos para la misma función.

### Sintaxis

```
VAR_SAMP | VARIANCE ( [ DISTINCT | ALL ] expression)
VAR_POP ( [ DISTINCT | ALL ] expression)
```

La expresión debe ser un tipo de datos entero, decimal o de punto flotante. Independientemente del tipo de datos de la expresión, el tipo de retorno de esta función es un número de doble precisión.

> **Note**
>
> Los resultados de estas funciones pueden variar entre clústeres de data warehouse, según la configuración del clúster en cada caso.

### Notas de uso

Cuando la varianza de la muestra (VARIANCE o VAR_SAMP) se calcula para una expresión que consta de un valor único, el resultado de la función es NULL no 0.

### Ejemplos

La siguiente consulta devuelve la varianza redondeada de muestra y población de la columna NUMTICKETS en la tabla LISTING.

```
select avg(numtickets),
round(var_samp(numtickets)) varsamp,
round(var_pop(numtickets)) varpop
from listing;

avg | varsamp | varpop
-----+---------+--------
10 |      54 |     54
(1 row)
```

La siguiente consulta ejecuta los mismos cálculos pero convierte los resultados a valores decimales.

```
select avg(numtickets),
cast(var_samp(numtickets) as dec(10,4)) varsamp,
cast(var_pop(numtickets) as dec(10,4)) varpop
from listing;
```

```
avg | varsamp | varpop
-----+---------+---------
10 | 53.6291 | 53.6288
(1 row)
```

# Funciones de matriz

A continuación, encontrará una descripción de las funciones de matriz de SQL que Amazon Redshift admite para acceder a las matrices y manipularlas.

Temas

- Función array
- función array_concat
- Función array_flatten
- Función get_array_length
- Función split_to_array
- función de submatriz

## Función array

Crea una matriz cuyo tipo de datos es SUPER.

Sintaxis

```
ARRAY( [ expr1 ] [ , expr2 [ , ... ] ] )
```

Argumento

expr1, expr2

Expresiones de cualquier tipo de datos de Amazon Redshift, excepto los tipos de fecha y hora, ya que Amazon Redshift no convierte los tipos de fecha y hora al tipo de datos SUPER. Los argumentos no tienen que ser del mismo tipo de datos.

Tipo de retorno

La función array devuelve el tipo de datos SUPER.

### Ejemplo

En los siguientes ejemplos, se muestra una matriz de valores numéricos y una matriz de diferentes tipos de datos.

```
--an array of numeric values
select array(1,50,null,100);
      array
------------------
 [1,50,null,100]
(1 row)

--an array of different data types
select array(1,'abc',true,3.14);
        array
----------------------
 [1,"abc",true,3.14]
(1 row)
```

## función array_concat

La función array_concat concatena dos matrices para crear una matriz que contiene todos los elementos de la primera matriz seguida de todos los elementos de la segunda matriz. Los dos argumentos deben ser matrices válidas.

### Sintaxis

```
array_concat( super_expr1,  super_expr2 )
```

### Argumentos

super_expr1

El valor que especifica la primera de las dos matrices que se van a concatenar.

super_expr2

El valor que especifica la segunda de las dos matrices que se van a concatenar.

### Tipo de retorno

La función array_concat devuelve un valor de datos SUPER.

### Ejemplo

En los siguientes ejemplos, se muestra la concatenación de dos matrices del mismo tipo y la concatenación de dos matrices de distintos tipos.

```
-- concatenating two arrays
SELECT ARRAY_CONCAT(ARRAY(10001,10002),ARRAY(10003,10004));
              array_concat
-------------------------------------
 [10001,10002,10003,10004]
(1 row)

-- concatenating two arrays of different types
SELECT ARRAY_CONCAT(ARRAY(10001,10002),ARRAY('ab','cd'));
          array_concat
------------------------------
 [10001,10002,"ab","cd"]
(1 row)
```

## Función array_flatten

Fusiona varias matrices en una sola matriz de tipo SUPER.

### Sintaxis

```
array_flatten( super_expr1,super_expr2,.. )
```

### Argumentos

super_expr1,super_expr2

Una expresión SUPER válida en forma de matriz.

### Tipo de retorno

La función array_flatten devuelve un valor de datos SUPER.

### Ejemplo

En el siguiente ejemplo, se muestra una función array_flatten.

```
SELECT ARRAY_FLATTEN(ARRAY(ARRAY(1,2,3,4),ARRAY(5,6,7,8),ARRAY(9,10)));
```

```
     array_flatten
-----------------------
 [1,2,3,4,5,6,7,8,9,10]
(1 row)
```

## Función get_array_length

Devuelve la longitud de la matriz especificada. La función GET_ARRAY_LENGTH devuelve la longitud de una matriz SUPER dada una ruta de objeto o matriz.

### Sintaxis

```
get_array_length( super_expr )
```

### Argumentos

super_expr

Una expresión SUPER válida en forma de matriz.

### Tipo de retorno

La función get_array_length devuelve un BIGINT.

### Ejemplo

En el siguiente ejemplo, se muestra una función get_array_length.

```
SELECT GET_ARRAY_LENGTH(ARRAY(1,2,3,4,5,6,7,8,9,10));
 get_array_length
----------------------
            10
(1 row)
```

## Función split_to_array

Utiliza un delimitador como parámetro opcional. Si ningún delimitador está presente, entonces el valor predeterminado es la coma.

### Sintaxis

```
split_to_array( string,delimiter )
```

Argumentos

string

La cadena de entrada que se dividirá.

delimiter

Un valor opcional en el que se dividirá la cadena de entrada. El valor predeterminado es una coma.

Tipo de retorno

La función split_to_array devuelve un valor de datos SUPER.

Ejemplo

En el siguiente ejemplo, se muestra una función split_to_array.

```
SELECT SPLIT_TO_ARRAY('12|345|6789', '|');
     split_to_array
------------------------
 ["12","345","6789"]
(1 row)
```

## función de submatriz

Manipula las matrices para devolver un subconjunto de las matrices de entrada.

Sintaxis

```
SUBARRAY( super_expr, start_position, length )
```

Argumentos

super_expr

Una expresión válida SUPER en forma de matriz.

start_position

La posición dentro de la matriz desde donde comienza la extracción, con punto de partida en la posición de índice 0. Con una posición negativa, se cuenta hacia atrás desde el final de la matriz.

longitud

La cantidad de elementos para extraer (la longitud de la subcadena).

Tipo de retorno

La función subarray devuelve un valor de datos SUPER.

Ejemplos

El siguiente es un ejemplo de una función subarray.

```
 SELECT SUBARRAY(ARRAY('a', 'b', 'c', 'd', 'e', 'f'), 2, 3);
   subarray
---------------
 ["c","d","e"]
(1 row)
```

## Funciones de agregación bit a bit

Las funciones de agregación bit a bit calculan operaciones de bits para realizar la agregación de columnas de enteros y columnas que se pueden convertir o redondear a valores enteros.

Temas

- Uso de NULL en agrupaciones bit a bit
- DISTINCT compatible con agregaciones bit a bit
- Información general de ejemplos de funciones bit a bit
- Función BIT_AND
- Función BIT_OR
- Función BOOL_AND
- Función BOOL_OR

### Uso de NULL en agrupaciones bit a bit

Cuando se aplica una función bit a bit a una columna que puede contener valores nulos, se eliminan todos los valores NULL antes de que se calcule el resultado de la función. Si ninguna fila califica para

la agregación, la función bit a bit devuelve NULL. El mismo comportamiento se aplica a funciones de agregación regulares. A continuación se muestra un ejemplo.

```
select sum(venueseats), bit_and(venueseats) from venue
where venueseats is null;

sum  | bit_and
------+---------
null |    null
(1 row)
```

## DISTINCT compatible con agregaciones bit a bit

Como otras funciones de agrupación, las funciones bit a bit son compatibles con la palabra clave DISTINCT.

No obstante, usar DISTINCT con estas funciones no influye en los resultados. La primera instancia de un valor es suficiente para satisfacer las operaciones bit a bit AND u OR. No importa si hay valores duplicados en la expresión que se evalúa.

Debido a que es probable que el procesamiento de DISTINCT genere algunos gastos de ejecución de consulta, se recomienda no usar DISTINCT con las funciones bit a bit.

## Información general de ejemplos de funciones bit a bit

A continuación, encontrará algunos ejemplos de información general que demuestran cómo trabajar con las funciones bit a bit. También puede encontrar ejemplos de código específicos con cada descripción de la función.

Los ejemplos de las funciones bit a bit se basan en la base de datos de muestra TICKIT. La tabla USERS de la base de datos de muestra TICKIT tiene varias columnas con valores booleanos que indican si cada usuario es conocido por tener tipos diferentes de eventos, como deportes, teatro, ópera, etc. Ejemplo:

```
select userid, username, lastname, city, state,
likesports, liketheatre
from users limit 10;

userid | username | lastname  |     city     | state | likesports | liketheatre
-------+----------+-----------+--------------+-------+------------+-------------
```

```
1 | JSG99FHE | Taylor    | Kent         | WA    | t          | t
9 | MSD36KVR | Watkins   | Port Orford  | MD    | t          | f
```

Suponga que el diseño de una versión nueva de la tabla USERS es diferente. En esta versión nueva, hay una única columna con valores enteros que define (en formato binario) ocho tipos de eventos que a cada usuario les gustan o les disgustan. En este diseño, cada posición de bit representa un tipo de evento. Un usuario al que le gustan los ocho tipos tiene los ocho bits establecidos en 1 (como en la primera fila de la siguiente tabla). Un usuario al que no le gusta ninguno de estos eventos tiene los ocho bits establecidos en 0 (vea la segunda fila). En la tercera fila que se encuentra a continuación, se representa a un usuario al que solo le gustan los deportes y el jazz.

| USER | DEPORTE | TEATRO | JAZZ | OPERA | ROCK | LAS VEGAS | BROADW | CLÁSICO |
|---|---|---|---|---|---|---|---|---|
| Usuario 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Usuario 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Usuario 3 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |

En la tabla de base de datos, estos valores binarios se pueden almacenar en una única columna denominada LIKES como valores enteros, como se muestra a continuación.

| Usuario | Valor binario | Valor almacenado (entero) |
|---|---|---|
| Usuario 1 | 11111111 | 255 |
| Usuario 2 | 00000000 | 0 |
| Usuario 3 | 10100000 | 160 |

## Función BIT_AND

La función BIT_AND ejecuta operaciones AND bit a bit en todos los valores de una única columna o expresión con valores enteros. Estas funciones agregan cada bit de cada valor binario que corresponde a cada valor entero en la expresión.

La función BIT_AND devuelve un resultado de `0` si ninguno de los bits se establece en 1 en todos los valores. Si uno o más bits se establece en 1 en todos los valores, la función devuelve un valor entero. Este entero es el número que corresponde al valor binario para esos bits.

Por ejemplo, una tabla tiene cuatro valores enteros en una columna: 3, 7, 10 y 22. Esos enteros están representados en formato binario de la siguiente manera:

| Entero | Valor binario |
|---|---|
| 3 | 11 |
| 7 | 111 |
| 10 | 1010 |
| 22 | 10110 |

Una operación BIT_AND en este conjunto de datos encuentra que todos los bits se establecen en 1 solo en la penúltima posición. El resultado es un valor binario de `00000010`, que representa el valor entero 2. Por lo tanto, la función BIT_AND devuelve 2.

Sintaxis

```
BIT_AND ( [DISTINCT | ALL] expression )
```

Argumentos

expression

La columna o expresión de destino sobre la que opera la función. Esta expresión debe tener un tipo de datos INT, INT2 o INT8. La función devuelve un tipo de datos INT, INT2 o INT8 equivalente.

DISTINCT | ALL

Con el argumento DISTINCT, la función elimina todos los valores duplicados para la expresión especificada antes de calcular el resultado. Con el argumento ALL, la función retiene todos los valores duplicados. El valor predeterminado es ALL. Para obtener más información, consulte DISTINCT compatible con agregaciones bit a bit.

### Ejemplos

Dado que esa información significativa de negocio se almacena en columnas con valores enteros, puede usar funciones bit a bit para extraer y acumular esa información. La siguiente consulta aplica la función BIT_AND a la columna LIKES en una tabla que se llama USERLIKES y agrupa los resultados según la columna CITY.

```
select city, bit_and(likes) from userlikes group by city
order by city;
city          | bit_and
--------------+---------
Los Angeles   |       0
Sacramento    |       0
San Francisco |       0
San Jose      |      64
Santa Barbara |     192
(5 rows)
```

Estos resultados se pueden interpretar de la siguiente manera:

- El valor entero `192` para Santa Barbara se traduce al valor binario `11000000`. En otras palabras, a todos los usuarios de esta ciudad les gustan deportes y teatro, pero no a todos les gusta otro tipo de evento.
- El valor entero `64` se traduce a `01000000`. Entonces, para los usuarios de San José, el único tipo de evento que les gusta a todos es el teatro.
- Los valores de `0` para las otras tres ciudades indican que no se comparten "gustos" entre todos los usuarios de esas ciudades.

## Función BIT_OR

La función BIT_OR ejecuta operaciones OR bit a bit en todos los valores de una única columna o expresión con valores enteros. Estas funciones agregan cada bit de cada valor binario que corresponde a cada valor entero en la expresión.

Por ejemplo, suponga que su tabla tiene cuatro valores enteros en una columna: 3, 7, 10 y 22. Esos valores enteros están representados en formato binario de la siguiente manera.

| Entero | Valor binario |
|---|---|
| 3 | 11 |
| 7 | 111 |
| 10 | 1010 |
| 22 | 10110 |

Si aplica la función BIT_OR al conjunto de valores enteros, la operación busca cualquier valor donde se haya encontrado un 1 en cada posición. En este caso, existe un valor de 1 en las últimas cinco posiciones para al menos uno de los valores, lo que da como resultado un valor binario de `00011111`; por lo tanto, la función devuelve `31` (o `16 + 8 + 4 + 2 + 1`).

Sintaxis

```
BIT_OR ( [DISTINCT | ALL] expression )
```

Argumentos

expression

La columna o expresión de destino sobre la que opera la función. Esta expresión debe tener un tipo de datos INT, INT2 o INT8. La función devuelve un tipo de datos INT, INT2 o INT8 equivalente.

DISTINCT | ALL

Con el argumento DISTINCT, la función elimina todos los valores duplicados para la expresión especificada antes de calcular el resultado. Con el argumento ALL, la función retiene todos los valores duplicados. El valor predeterminado es ALL. Para obtener más información, consulte DISTINCT compatible con agregaciones bit a bit.

Ejemplo

La siguiente consulta aplica la función BIT_OR a la columna LIKES en una tabla que se llama USERLIKES y agrupa los resultados según la columna CITY.

```
select city, bit_or(likes) from userlikes group by city
order by city;
city          | bit_or
--------------+--------
Los Angeles   |    127
Sacramento    |    255
San Francisco |    255
San Jose      |    255
Santa Barbara |    255
(5 rows)
```

Para cuatro de las ciudades indicadas, todos los tipos de eventos tienen el gusto de, al menos, un usuario (`255=11111111`). Para Los Angeles, todos los tipos de eventos, excepto deportes, tienen el gusto de al menos un usuario (`127=01111111`).

## Función BOOL_AND

La función BOOL_AND funciona en una sola columna o expresión con valores booleanos o enteros. Esta función aplica una lógica similar a las funciones BIT_AND y BIT_OR. Para esta función, el tipo de retorno es un valor booleano (`true` o `false`).

Si todos los valores de un conjunto son verdaderos, la función BOOL_AND devuelve `true` (`t`). Si todo valor es falso, la función devuelve `false` (`f`).

### Sintaxis

```
BOOL_AND ( [DISTINCT | ALL] expression )
```

### Argumentos

expression

La columna o expresión de destino sobre la que opera la función. Esta expresión debe tener un tipo de datos booleano o entero. El tipo de retorno de la función es booleano.

DISTINCT | ALL

Con el argumento DISTINCT, la función elimina todos los valores duplicados para la expresión especificada antes de calcular el resultado. Con el argumento ALL, la función retiene todos los valores duplicados. El valor predeterminado es ALL. Para obtener más información, consulte DISTINCT compatible con agregaciones bit a bit.

### Ejemplos

Puede utilizar funciones booleanas con expresiones booleanas o expresiones enteras. Por ejemplo, la siguiente consulta devuelve resultados de la tabla estándar USERS en la base de datos TICKIT, que tiene varias columnas con valores booleanos.

La función BOOL_AND devuelve `false` para las cinco filas. A no todos los usuarios en cada uno de esos estados les gusta deportes.

```
select state, bool_and(likesports) from users
group by state order by state limit 5;

state | bool_and
------+---------
AB    | f
AK    | f
AL    | f
AZ    | f
BC    | f
(5 rows)
```

## Función BOOL_OR

La función BOOL_OR funciona en una única columna o expresión booleana o entera. Esta función aplica una lógica similar a las funciones BIT_AND y BIT_OR. Para esta función, el tipo de retorno es un valor booleano (`true`, `false` o `NULL`).

Si uno o varios valores en un conjunto es `true`, la función BOOL_OR devuelve `true` (`t`). Si todos los valores de un conjunto son `false`, la función devuelve `false` (`f`). Se puede devolver NULL si se desconoce el valor.

### Sintaxis

```
BOOL_OR ( [DISTINCT | ALL] expression )
```

### Argumentos

expression

La columna o expresión de destino sobre la que opera la función. Esta expresión debe tener un tipo de datos booleano o entero. El tipo de retorno de la función es booleano.

DISTINCT | ALL

Con el argumento DISTINCT, la función elimina todos los valores duplicados para la expresión especificada antes de calcular el resultado. Con el argumento ALL, la función retiene todos los valores duplicados. El valor predeterminado es ALL. Consulte DISTINCT compatible con agregaciones bit a bit.

Ejemplos

Puede utilizar las funciones booleanas con expresiones booleanas o expresiones enteras. Por ejemplo, la siguiente consulta devuelve resultados de la tabla estándar USERS en la base de datos TICKIT, que tiene varias columnas con valores booleanos.

La función BOOL_OR devuelve `true` para las cinco filas. A al menos un usuario en cada uno de esos estados les gusta deportes.

```
select state, bool_or(likesports) from users
group by state order by state limit 5;

state | bool_or
------+--------
AB    | t
AK    | t
AL    | t
AZ    | t
BC    | t
(5 rows)
```

El ejemplo siguiente devuelve NULL.

```
SELECT BOOL_OR(NULL = '123')
               bool_or
------
NULL
```

## Expresiones condicionales

Temas

- Expresión condicional CASE
- Función DECODE

- Funciones GREATEST y LEAST
- Funciones NVL y COALESCE
- Función NVL2
- Función NULLIF

Amazon Redshift admite algunas expresiones condicionales que son extensiones del estándar SQL.

## Expresión condicional CASE

La expresión CASE es una expresión condicional similar a las instrucciones if/then/else que se encuentran en otros lenguajes. CASE se utiliza para especificar un resultado cuando hay condiciones múltiples. Utilice CASE cuando una expresión SQL sea válida, como en un comando SELECT.

Existen dos tipos de expresiones CASE: simple y buscada.

- En expresiones CASE simples, una expresión se compara con un valor. Cuando hay una coincidencia, se aplica la acción especificada en la cláusula THEN. Si no se encuentra coincidencia, se aplica la acción en la cláusula ELSE.
- En las expresiones CASE buscadas, cada CASE se evalúa según una expresión booleana, y la instrucción CASE devuelve el primer CASE que coincida. Si no hay ninguna coincidencia entre las cláusulas WHEN, se devuelve la acción en la cláusula ELSE.

### Sintaxis

Instrucción CASE simple utilizada para hacer coincidir condiciones:

```
CASE expression
  WHEN value THEN result
  [WHEN...]
  [ELSE result]
END
```

Instrucción CASE buscada utilizada para evaluar cada condición:

```
CASE
  WHEN condition THEN result
  [WHEN ...]
  [ELSE result]
```

```
END
```

### Argumentos

expresión

Un nombre de columna o cualquier expresión válida.

value

Valor con el que se compara la expresión, como una constante numérica o una cadena de caracteres.

result

El valor destino o la expresión que se devuelve cuando se evalúa una expresión o una condición booleana. Los tipos de datos de todas las expresiones de resultado deben poder convertirse a un único tipo de salida.

condition

Expresión booleana que se evalúa como true o false. Si el argumento condition es verdadero, el valor de la expresión CASE es el resultado que sigue a la condición y el resto de la expresión CASE no se procesa. Si el argumento condition es falso, se evalúan las cláusulas WHEN subsiguientes. Si ningún resultado de la condición WHEN es verdadero, el valor de la expresión CASE será el resultado de la cláusula ELSE. Si se omite la cláusula ELSE y ninguna condición es verdadera, el resultado será nulo.

### Ejemplos

En los siguientes ejemplos se utilizan las tablas VENUE y SALES de los datos de muestra de TICKIT. Para obtener más información, consulte Base de datos de muestra.

Use una expresión CASE simple para reemplazar `New York City` por `Big Apple` en una consulta de la tabla VENUE. Reemplace todos los demás nombres de ciudad por `other`.

```
select venuecity,
  case venuecity
    when 'New York City'
    then 'Big Apple' else 'other'
  end
from venue
order by venueid desc;
```

```
venuecity        |   case
-----------------+-----------
Los Angeles      | other
New York City    | Big Apple
San Francisco    | other
Baltimore        | other
...
```

Utilice una expresión CASE buscada para asignar números de grupo según el valor PRICEPAID para ventas de tickets individuales:

```
select pricepaid,
  case when pricepaid <10000 then 'group 1'
    when pricepaid >10000 then 'group 2'
    else 'group 3'
  end
from sales
order by 1 desc;

pricepaid |  case
----------+---------
12624     | group 2
10000     | group 3
10000     | group 3
9996      | group 1
9988      | group 1
...
```

## Función DECODE

Una expresión DECODE reemplaza un valor específico por otro valor específico o por un valor predeterminado, según el resultado de una condición de igualdad. Esta operación equivale a la operación de una expresión CASE simple o una instrucción IF-THEN-ELSE.

### Sintaxis

```
DECODE ( expression, search, result [, search, result ]... [ ,default ] )
```

Este tipo de expresión es útil para reemplazar abreviaciones o códigos almacenados en tablas con valores comerciales significativos que se necesitan para los informes.

### Parámetros

expression

El origen del valor que desea comparar, como una columna en una tabla.

search

El valor destino con el que se compara la expresión origen, como un valor numérico o una cadena de caracteres. La expresión buscar debe tomar un único valor fijo. No puede especificar una expresión que tome un rango de valores, como `age between 20 and 29`; debe especificar pares separados de búsqueda-resultado para cada valor que desea reemplazar.

Los tipos de datos para todas las instancias de la expresión buscar deben ser iguales o compatibles. Los parámetros expression (expresión) y search (buscar) también deben ser compatibles.

result

El valor de reemplazo que devuelve la consulta cuando la expresión coincide con el valor de búsqueda. Debe incluir al menos un par búsqueda-resultado en la expresión DECODE.

Los tipos de datos para todas las instancias de la expresión resultado deben ser iguales o compatibles. Los parámetros result (resultado) y default (predeterminado) también deben ser compatibles.

predeterminada

Un valor predeterminado opcional que se usa para casos en los que falla la condición de búsqueda. Si no especifica un valor predeterminado, la expresión DECODE devuelve NULL.

### Notas de uso

Si el valor de expression y el valor search son los dos NULL, el resultado DECODE es el valor result correspondiente. Para ver una ilustración de este uso de la función, consulte la sección Ejemplos.

Cuando se usa de esta manera, DECODE es similar a Función NVL2, pero hay algunas diferencias. Para obtener una descripción de estas diferencias, consulte las notas de uso de NVL2.

### Ejemplos

Cuando el valor `2008-06-01` existe en la columna caldate de datatable, el siguiente ejemplo lo reemplaza por `June 1st, 2008`. El ejemplo reemplaza todos los demás valores caldate por NULL.

```
select decode(caldate, '2008-06-01', 'June 1st, 2008')
from datetable where month='JUN' order by caldate;

case
----------------
June 1st, 2008

...
(30 rows)
```

En el siguiente ejemplo, se utiliza una expresión DECODE para convertir a nombres completos las cinco columnas CATNAME abreviadas en la tabla CATEGORY y convertir los demás valores de la columna en Unknown.

```
select catid, decode(catname,
'NHL', 'National Hockey League',
'MLB', 'Major League Baseball',
'MLS', 'Major League Soccer',
'NFL', 'National Football League',
'NBA', 'National Basketball Association',
'Unknown')
from category
order by catid;

catid  | case
-------+---------------------------------
1      | Major League Baseball
2      | National Hockey League
3      | National Football League
4      | National Basketball Association
5      | Major League Soccer
6      | Unknown
7      | Unknown
8      | Unknown
9      | Unknown
10     | Unknown
11     | Unknown
(11 rows)
```

Se usa una expresión DECODE para encontrar lugares en Colorado y Nevada con NULL en la columna VENUESEATS; se convierten los NULL a ceros. Si la columna VENUESEATS no es NULL, devuelve 1 como resultado.

```
select venuename, venuestate, decode(venueseats,null,0,1)
from venue
where venuestate in('NV','CO')
order by 2,3,1;

venuename               | venuestate     | case
----------------------------+---------------+-----------
Coors Field                   | CO        |   1
Dick's Sporting Goods Park    | CO        |   1
Ellie Caulkins Opera House    | CO        |   1
INVESCO Field        | CO        |   1
Pepsi Center        | CO        |   1
Ballys Hotel        | NV        |   0
Bellagio Hotel                | NV        |   0
Caesars Palace                | NV        |   0
Harrahs Hotel                 | NV        |   0
Hilton Hotel                  | NV        |   0
...
(20 rows)
```

## Funciones GREATEST y LEAST

Devuelve el valor más grande o el más pequeño de una lista de cualquier cantidad de expresiones.

### Sintaxis

```
GREATEST (value [, ...])
LEAST (value [, ...])
```

### Parámetros

expression_list

Una lista de expresiones separada por comas, como la columna nombres. Las expresiones deben ser todas convertibles a un tipo común de datos. Se ignoran los valores NULL en la lista. Si todas las expresiones toman el valor NULL, el resultado es NULL.

### Devuelve

Devuelve el valor máximo (para GREATEST) o el mínimo (para LEAST) de la lista de expresiones proporcionada.

### Ejemplo

El siguiente ejemplo devuelve el valor más alto alfabéticamente para `firstname` o `lastname`.

```
select firstname, lastname, greatest(firstname,lastname) from users
where userid < 10
order by 3;

 firstname | lastname  | greatest
-----------+-----------+-----------
 Lars      | Ratliff   | Ratliff
 Reagan    | Hodge     | Reagan
 Colton    | Roy       | Roy
 Barry     | Roy       | Roy
 Tamekah   | Juarez    | Tamekah
 Rafael    | Taylor    | Taylor
 Victor    | Hernandez | Victor
 Vladimir  | Humphrey  | Vladimir
 Mufutau   | Watkins   | Watkins
(9 rows)
```

## Funciones NVL y COALESCE

Devuelve el valor de la primera expresión que no es nula en una serie de expresiones. Cuando se encuentra un valor que no es nulo, las expresiones restantes de la lista no se evalúan.

NVL es idéntica a COALESCE. Son sinónimos. En este tema se explica la sintaxis y se incluyen ejemplos de ambas funciones.

### Sintaxis

```
NVL( expression, expression, ... )
```

La sintaxis de COALESCE es la misma:

```
COALESCE( expression, expression, ... )
```

Si todas las expresiones son nulas, el resultado es nulo.

Estas funciones son útiles cuando se desea devolver un valor secundario si falta un valor primario o es nulo. Por ejemplo, una consulta puede devolver el primero de los tres números de teléfono

disponibles: móvil, fijo o trabajo. El orden de las expresiones de la función determina el orden de evaluación.

Argumentos

expresión

Una expresión, como un nombre de columna, que evalúa estados nulos.

Tipo de retorno

Amazon Redshift determina el tipo de datos del valor devuelto según las expresiones de entrada. Si los tipos de datos de las expresiones de entrada no tienen un tipo común, se devuelve un error.

Ejemplos

Si la lista contiene expresiones de enteros, la función devuelve un entero.

```
SELECT COALESCE(NULL, 12, NULL);

coalesce
--------------
12
```

Este ejemplo, que es igual al anterior, excepto que usa NVL, devuelve el mismo resultado.

```
SELECT NVL(NULL, 12, NULL);

coalesce
--------------
12
```

En el siguiente ejemplo, se devuelve un tipo de cadena.

```
SELECT COALESCE(NULL, 'Amazon Redshift', NULL);

coalesce
--------------
Amazon Redshift
```

En el siguiente ejemplo, se produce un error porque los tipos de datos varían en la lista de expresiones. En este caso, hay un tipo de cadena y un tipo de número en la lista.

```
SELECT COALESCE(NULL, 'Amazon Redshift', 12);
ERROR: invalid input syntax for integer: "Amazon Redshift"
```

Para este ejemplo, se crea una tabla con las columnas START_DATE y END_DATE, se insertan filas que incluyan valores nulos y, después, se aplica una expresión NVL a las dos columnas.

```
create table datetable (start_date date, end_date date);
insert into datetable values ('2008-06-01','2008-12-31');
insert into datetable values (null,'2008-12-31');
insert into datetable values ('2008-12-31',null);
```

```
select nvl(start_date, end_date)
from datetable
order by 1;

coalesce
------------
2008-06-01
2008-12-31
2008-12-31
```

El nombre predeterminado de la columna para una expresión NVL es COALESCE. La siguiente consulta devuelve los mismos resultados:

```
select coalesce(start_date, end_date)
from datetable
order by 1;
```

En las siguientes consultas de ejemplo, se crea una tabla con información de ejemplo sobre reservas de hotel y se insertan varias filas. Algunos registros contienen valores nulos.

```
create table booking_info (booking_id int, booking_code character(8), check_in date,
 check_out date, funds_collected numeric(12,2));
```

Inserte los siguientes datos de ejemplo. Algunos registros no tienen fecha `check_out` ni cantidad `funds_collected`.

```
insert into booking_info values (1, 'OCEAN_WV', '2023-02-01','2023-02-03',100.00);
insert into booking_info values (2, 'OCEAN_WV', '2023-04-22','2023-04-26',120.00);
insert into booking_info values (3, 'DSRT_SUN', '2023-03-13','2023-03-16',125.00);
```

```
insert into booking_info values (4, 'DSRT_SUN', '2023-06-01','2023-06-03',140.00);
insert into booking_info values (5, 'DSRT_SUN', '2023-07-10',null,null);
insert into booking_info values (6, 'OCEAN_WV', '2023-08-15',null,null);
```

La siguiente consulta devuelve una lista de fechas. Si la fecha de `check_out` no está disponible, se muestra la fecha de `check_in`.

```
select coalesce(check_out, check_in)
from booking_info
order by booking_id;
```

Los resultados son los siguientes. Tenga en cuenta que los dos últimos registros muestran la fecha de `check_in`.

```
coalesce
------------
2023-02-03
2023-04-26
2023-03-16
2023-06-03
2023-07-10
2023-08-15
```

Si espera que una consulta devuelva valores nulos para determinadas funciones o columnas, puede usar una expresión NVL para reemplazar los valores nulos por algún otro valor. Por ejemplo, funciones de agregación, como SUM, devolverían valores nulos en lugar de ceros cuando no tengan filas para evaluar. Puede usar una expresión NVL para reemplazar esos valores nulos por `700.0`. En lugar de `485`, el resultado de sumar `funds_collected` es `1885` porque dos filas que tienen un valor nulo se sustituyen por `700`.

```
select sum(nvl(funds_collected, 700.0)) as sumresult from booking_info;

sumresult
------
 1885
```

## Función NVL2

Devuelve uno de los dos valores, en función de si una expresión especificada toma un valor NULL o NOT NULL.

### Sintaxis

```
NVL2 ( expression, not_null_return_value, null_return_value )
```

### Argumentos

expresión

Una expresión, como un nombre de columna, que evalúa estados nulos.

not_null_return_value

El valor devuelto si la expression (expresión) toma un valor NOT NULL. El valor not_null_return_value debe tener los mismos tipos de datos que expression (expresión) o ser convertible implícitamente a ese tipo de datos.

null_return_value

El valor de retorno si expression (expresión) toma un valor NULL. El valor null_return_value debe tener los mismos tipos de datos que expression (expresión) o ser convertible implícitamente a ese tipo de datos.

### Tipo de retorno

El tipo de retorno NVL2 se determina de la siguiente manera:

- Si alguno de los valores not_null_return_value o null_return_value es nulo, se devuelve el tipo de datos de la expresión no nula.

Si ninguno de los valores not_null_return_value y null_return_value es nulo:

- Si los valores not_null_return_value y null_return_value tienen el mismo tipo de datos, se devuelve ese tipo de datos.
- Si los valores not_null_return_value y null_return_value tienen tipos de datos numéricos diferentes, se devuelve el tipo de dato numérico compatible que sea menor.
- Si los valores not_null_return_value y null_return_value tienen tipos de datos de fecha y hora diferentes, se devuelve un tipo de dato de marca temporal.
- Si los valores not_null_return_value y null_return_value tienen tipos de datos de caracteres diferentes, se devuelve el tipo de dato de not_null_return_value.

- Si los valores not_null_return_value y null_return_value tienen tipos de datos numéricos y no numéricos mezclados, se devuelve el tipo de dato de not_null_return_value.

> **Important**
>
> En los últimos dos casos en los que se devuelve el tipo de dato not_null_return_value, null_return_value está vinculado implícitamente a ese tipo de dato. Si los tipos de datos son incompatibles, la función falla.

### Notas de uso

Función DECODE se puede utilizar de manera similar a NVL2 cuando los parámetros expression y search son los dos nulos. La diferencia es que, para DECODE, el valor devuelto será el valor y el tipo de datos del parámetro result. Por el contrario, en el caso de NVL2, el valor devuelto tendrá el valor del parámetro not_null_return_value o del parámetro null_return_value, el que haya seleccionado la función, pero tendrá el tipo de datos de not_null_return_value.

Por ejemplo, si se asume que column1 es NULL, las siguientes consultas devolverán el mismo valor. No obstante, el tipo de datos de valor de retorno DECODE será INTEGER y el tipo de datos del valor de retorno NVL2 será VARCHAR.

```
select decode(column1, null, 1234, '2345');
select nvl2(column1, '2345', 1234);
```

### Ejemplo

En el siguiente ejemplo, se modifican algunos datos de muestra y, luego, se evalúan dos campos para proporcionar la información de contacto adecuada para los usuarios:

```
update users set email = null where firstname = 'Aphrodite' and lastname = 'Acevedo';

select (firstname + ' ' + lastname) as name,
nvl2(email, email, phone) AS contact_info
from users
where state = 'WA'
and lastname  like 'A%'
order by lastname, firstname;

name        contact_info
```

```
--------------------+--------------------------------------------
Aphrodite Acevedo (906) 632-4407
Caldwell Acevedo  Nunc.sollicitudin@Duisac.ca
Quinn Adams    vel@adipiscingligulaAenean.com
Kamal Aguilar  quis@vulputaterisusa.com
Samson Alexander  hendrerit.neque@indolorFusce.ca
Hall Alford    ac.mattis@vitaediamProin.edu
Lane Allen     et.netus@risusDonec.org
Xander Allison    ac.facilisis.facilisis@Infaucibus.com
Amaya Alvarado    dui.nec.tempus@eudui.edu
Vera Alvarez   at.arcu.Vestibulum@pellentesque.edu
Yetta Anthony  enim.sit@risus.org
Violet Arnold  ad.litora@at.com
August Ashley  consectetuer.euismod@Phasellus.com
Karyn Austin   ipsum.primis.in@Maurisblanditenim.org
Lucas Ayers    at@elitpretiumet.com
```

## Función NULLIF

### Sintaxis

La expresión NULLIF compara dos argumentos y devuelve un valor nulo si los argumentos son iguales. Si no son iguales, se devuelve el primer argumento. Esta expresión realiza lo contrario a lo que realiza la expresión NVL o COALESCE.

```
NULLIF ( expression1, expression2 )
```

### Argumentos

expresión1, expresión2

Las columnas o expresiones de destino que se comparan. El tipo de retorno es el mismo que el tipo de la primera expresión. El nombre predeterminado de la columna del resultado NULLIF es el nombre de columna de la primera expresión.

### Ejemplos

En el ejemplo siguiente, la consulta devuelve la cadena `first` porque los argumentos no son iguales.

```
SELECT NULLIF('first', 'second');
```

```
case
-------
first
```

En el ejemplo siguiente, la consulta devuelve NULL porque los argumentos literales de la cadena son iguales.

```
SELECT NULLIF('first', 'first');

case
-------
NULL
```

En el ejemplo siguiente, la consulta devuelve 1 porque los argumentos de enteros no son iguales.

```
SELECT NULLIF(1, 2);

case
-------
1
```

En el ejemplo siguiente, la consulta devuelve NULL porque los argumentos de enteros son iguales.

```
SELECT NULLIF(1, 1);

case
-------
NULL
```

En el siguiente ejemplo, la consulta devuelve valores nulos cuando los valores LISTID y SALESID coinciden:

```
select nullif(listid,salesid), salesid
from sales where salesid<10 order by 1, 2 desc;

listid  | salesid
--------+---------
     4  |       2
     5  |       4
     5  |       3
     6  |       5
     10 |       9
```

```
     10 |       8
     10 |       7
     10 |       6
        |       1
(9 rows)
```

Puede usar NULLIF para asegurarse de que las cadenas vacías se devuelvan siempre como valores nulos. En el siguiente ejemplo, la expresión NULLIF devuelve un valor nulo o una cadena que tiene al menos un carácter.

```
insert into category
values(0,'','Special','Special');

select nullif(catgroup,'') from category
where catdesc='Special';

catgroup
----------
null
(1 row)
```

NULLIF ignora los espacios en blanco a la derecha. Si una cadena no está vacía porque contiene espacios en blanco, NULLIF seguirá devolviendo un valor nulo:

```
create table nulliftest(c1 char(2), c2 char(2));

insert into nulliftest values ('a','a ');

insert into nulliftest values ('b','b');


select nullif(c1,c2) from nulliftest;
c1
------
null
null
(2 rows)
```

## Funciones de formato de tipo de datos

Temas

- Función CAST
- Función CONVERT
- TEXT_TO_INT_ALT
- TEXT_TO_NUMERIC_ALT
- TO_CHAR
- Función TO_DATE
- TO_NUMBER
- Función TRY_CAST
- Cadenas de formatos de fecha y hora
- Cadenas de formatos numéricos
- Caracteres de formato de estilo Teradata para datos numéricos

Las funciones de formato de tipo de datos ofrecen una manera fácil de convertir valores de un tipo de datos a otro. En cada una de estas funciones, el primer argumento siempre es el valor al que se va a dar formato, mientras que el segundo argumento contiene la plantilla del formato nuevo. Amazon Redshift admite varias funciones para dar formato con los tipos de datos.

## Función CAST

La función CAST convierte un tipo de datos en otro tipo compatible. Por ejemplo, puede convertir una cadena en una fecha o un tipo numérico en una cadena. CAST realiza una conversión en tiempo de ejecución, lo que significa que la conversión no cambia el tipo de datos de un valor en una tabla de origen. Solo cambia en el contexto de la consulta.

La función CAST es muy similar a the section called “CONVERT”, ya que ambas convierten un tipo de datos a otro, pero tienen nombres distintos.

Algunas tipos de datos requieren una conversión explícita a otros tipos de datos utilizando las funciones CAST o CONVERT. Otros tipos de datos se pueden convertir implícitamente, como parte de otro comando, sin utilizar CAST ni CONVERT. Consulte Conversión y compatibilidad de tipos.

### Sintaxis

Utilice cualquiera de estas dos formas sintácticas equivalentes para convertir expresiones de un tipo de datos a otro.

```
CAST ( expression AS type )
```

```
expression :: type
```

Argumentos

expresión

Una expresión que toma el valor de uno o más valores, como un nombre de columna o un literal. La conversión de valores nulos devuelve valores nulos. La expresión no puede tener cadenas en blanco ni vacías.

type

Uno de los Tipos de datos admitidos.

Tipo de retorno

CAST devuelve el tipo de datos especificado por el argumento type.

> **Note**
>
> Amazon Redshift devuelve un error si intenta realizar una conversión problemática, como una conversión DECIMAL que pierde precisión, como la siguiente:
>
> ```
> select 123.456::decimal(2,1);
> ```
>
> o una conversión a un valor de INTEGER que genera un desbordamiento:
>
> ```
> select 12345678::smallint;
> ```

Ejemplos

Algunos de los ejemplos utilizan la base de datos de muestra TICKIT. Para obtener más información acerca de cómo configurar datos de ejemplo, consulte Cargar datos.

Las siguientes dos consultas son equivalentes. Ambas convierten un valor decimal en uno entero:

```
select cast(pricepaid as integer)
from sales where salesid=100;
```

```
pricepaid
-----------
162
(1 row)
```

```
select pricepaid::integer
from sales where salesid=100;

pricepaid
-----------
162
(1 row)
```

Lo siguiente produce un resultado similar. No requiere datos de muestra para ejecutarse:

```
select cast(162.00 as integer) as pricepaid;

pricepaid
-----------
162
(1 row)
```

En este ejemplo, los valores de una columna de marca temporal se convierten en fechas, lo que elimina la hora de cada resultado:

```
select cast(saletime as date), salesid
from sales order by salesid limit 10;

 saletime  | salesid
-----------+---------
2008-02-18 |       1
2008-06-06 |       2
2008-06-06 |       3
2008-06-09 |       4
2008-08-31 |       5
2008-07-16 |       6
2008-06-26 |       7
2008-07-10 |       8
2008-07-22 |       9
2008-08-06 |      10
(10 rows)
```

Si no utilizara CAST como se ilustra en el ejemplo anterior, los resultados incluirían la hora: 2008-02-18 02:36:48.

La siguiente consulta convierte los datos de caracteres variables en una fecha. No requiere datos de muestra para ejecutarse.

```
select cast('2008-02-18 02:36:48' as date) as mysaletime;

mysaletime
--------------------
2008-02-18
(1 row)
```

En este ejemplo, los valores en una columna de fecha se convierten en marcas temporales:

```
select cast(caldate as timestamp), dateid
from date order by dateid limit 10;

      caldate       | dateid
--------------------+--------
2008-01-01 00:00:00 |   1827
2008-01-02 00:00:00 |   1828
2008-01-03 00:00:00 |   1829
2008-01-04 00:00:00 |   1830
2008-01-05 00:00:00 |   1831
2008-01-06 00:00:00 |   1832
2008-01-07 00:00:00 |   1833
2008-01-08 00:00:00 |   1834
2008-01-09 00:00:00 |   1835
2008-01-10 00:00:00 |   1836
(10 rows)
```

En un caso como el del ejemplo anterior, puede obtener control adicional sobre el formato de salida si utiliza TO_CHAR.

En este ejemplo, un valor entero se convierte en una cadena de caracteres:

```
select cast(2008 as char(4));

bpchar
--------
2008
```

En este ejemplo, un valor DECIMAL (6,3) se convierte en un valor DECIMAL (4,1):

```
select cast(109.652 as decimal(4,1));

numeric
---------
109.7
```

Este ejemplo muestra una expresión más compleja. La columna PRICEPAID (una columna DECIMAL(8,2)) de la tabla SALES se convierte en una columna DECIMAL(38,2) y los valores se multiplican por 100000000000000000000:

```
select salesid, pricepaid::decimal(38,2)*100000000000000000000
as value from sales where salesid<10 order by salesid;


 salesid |           value
---------+----------------------------
       1 | 728000000000000000000000.00
       2 |  760000000000000000000000.00
       3 | 350000000000000000000000.00
       4 | 175000000000000000000000.00
       5 | 154000000000000000000000.00
       6 | 394000000000000000000000.00
       7 | 788000000000000000000000.00
       8 | 197000000000000000000000.00
       9 | 591000000000000000000000.00
(9 rows)
```

> **Note**
>
> No puede realizar una operación de CAST o CONVERTER en un tipo de datos `GEOMETRY` para cambiarlo a otro tipo de datos. No obstante, puede proporcionar una representación hexadecimal de una cadena literal en un formato extendido de binario conocido (EWKB) como entrada a las funciones que aceptan un argumento `GEOMETRY`. Por ejemplo, la función `ST_AsText` siguiente espera un tipo de datos `GEOMETRY`.
>
> ```
> SELECT ST_AsText('010100000000000000000001C400000000000002040');
> ```
>
> ```
> st_astext
> ```

```
------------
 POINT(7 8)
```

También puede especificar explícitamente el tipo de datos GEOMETRY.

```
SELECT ST_AsText('010100000000000000000014400000000000001840'::geometry);
```

```
st_astext
------------
 POINT(5 6)
```

## Función CONVERT

Al igual que la función CAST, la función CONVERT convierte un tipo de datos en otro tipo compatible. Por ejemplo, puede convertir una cadena en una fecha o un tipo numérico en una cadena. CONVERT realiza una conversión en tiempo de ejecución, lo que significa que la conversión no cambia el tipo de datos de un valor en una tabla de origen. Solo cambia en el contexto de la consulta.

Determinados tipos de datos requieren una conversión explícita a otros tipos de datos con la función CONVERT. Otros tipos de datos se pueden convertir implícitamente, como parte de otro comando, sin utilizar CAST ni CONVERT. Consulte Conversión y compatibilidad de tipos.

Sintaxis

```
CONVERT ( type, expression )
```

Argumentos

type

Uno de los Tipos de datos admitidos.

expresión

Una expresión que toma el valor de uno o más valores, como un nombre de columna o un literal. La conversión de valores nulos devuelve valores nulos. La expresión no puede tener cadenas en blanco ni vacías.

Tipo de retorno

CONVERT devuelve el tipo de datos especificado por el argumento type.

> **Note**
>
> Amazon Redshift devuelve un error si intenta realizar una conversión problemática, como una conversión DECIMAL que pierde precisión, como la siguiente:
>
> ```
> SELECT CONVERT(decimal(2,1), 123.456);
> ```
>
> o una conversión a un valor de INTEGER que genera un desbordamiento:
>
> ```
> SELECT CONVERT(smallint, 12345678);
> ```

Ejemplos

Algunos de los ejemplos utilizan la base de datos de muestra TICKIT. Para obtener más información acerca de cómo configurar datos de ejemplo, consulte Cargar datos.

En la siguiente consulta, se utiliza la función CONVERT para convertir una columna de decimales en enteros

```
SELECT CONVERT(integer, pricepaid)
FROM sales WHERE salesid=100;
```

En este ejemplo, se convierte un entero en una cadena de caracteres.

```
SELECT CONVERT(char(4), 2008);
```

En este ejemplo, la fecha y la hora actuales se convierten en un tipo de datos de carácter variable:

```
SELECT CONVERT(VARCHAR(30), GETDATE());

getdate
---------
2023-02-02 04:31:16
```

En este ejemplo, se convierte la columna saletime solo en la hora y se eliminan las fechas de cada fila.

```
SELECT CONVERT(time, saletime), salesid
FROM sales order by salesid limit 10;
```

Para obtener información sobre la conversión de una marca temporal de una zona horaria a otra, consulte Función CONVERT_TIMEZONE. Para obtener más funciones de fecha y hora, consulte Funciones de fecha y hora.

En el siguiente ejemplo, se convierten datos de caracteres variable en un objeto datetime.

```
SELECT CONVERT(datetime, '2008-02-18 02:36:48') as mysaletime;
```

> **Note**
>
> No puede realizar una operación de CAST o CONVERTER en un tipo de datos GEOMETRY para cambiarlo a otro tipo de datos. No obstante, puede proporcionar una representación hexadecimal de una cadena literal en un formato extendido de binario conocido (EWKB) como entrada a las funciones que aceptan un argumento GEOMETRY. Por ejemplo, la función ST_AsText siguiente espera un tipo de datos GEOMETRY.
>
> ```
> SELECT ST_AsText('010100000000000000000001C40000000000002040');
> ```
>
> ```
> st_astext
> ------------
>  POINT(7 8)
> ```
>
> También puede especificar explícitamente el tipo de datos GEOMETRY.
>
> ```
> SELECT ST_AsText('010100000000000000000014400000000000001840'::geometry);
> ```
>
> ```
> st_astext
> ------------
>  POINT(5 6)
> ```

# TEXT_TO_INT_ALT

TEXT_TO_INT_ALT convierte una cadena de caracteres en un valor entero utilizando el formato de estilo Teradata. Se truncan los dígitos de fracción en el resultado.

## Sintaxis

```
TEXT_TO_INT_ALT (expression [ , 'format'])
```

## Argumentos

expresión

Una expresión que produce uno o más valores CHAR o VARCHAR, como un nombre de columna o un literal de cadena. La conversión de valores nulos devuelve valores nulos. La función convierte las cadenas en blanco o vacías a 0.

format

Un literal de cadena que define el formato de la expresión de entrada. Para obtener más información acerca de los caracteres de formato que se pueden especificar, consulte Caracteres de formato de estilo Teradata para datos numéricos.

## Tipo de retorno

TEXT_TO_INT_ALT devuelve un valor INTEGER.

Se trunca la parte de fracción del resultado de la conversión.

Amazon Redshift devuelve un error si la conversión a la frase de formato que se especifica no consigue resultados correctos.

## Ejemplos

En el siguiente ejemplo, se convierte la cadena de expresión de entrada '123-' al valor entero -123.

```
select text_to_int_alt('123-');
```

```
text_to_int_alt
----------
      -123
```

En el siguiente ejemplo, se convierte la cadena de expresión de entrada '2147483647+' al valor entero 2147483647.

```
select text_to_int_alt('2147483647+');
```

```
text_to_int_alt
----------
2147483647
```

En el siguiente ejemplo, se convierte la cadena de expresión de entrada exponencial '-123E-2' al valor entero -1.

```
select text_to_int_alt('-123E-2');
```

```
text_to_int_alt
----------
        -1
```

En el siguiente ejemplo, se convierte la cadena de expresión de entrada '2147483647+' al valor entero 2147483647.

```
select text_to_int_alt('2147483647+');
```

```
text_to_int_alt
----------
2147483647
```

En el siguiente ejemplo, se convierte la cadena de expresión de entrada '123{' con la frase de formato '999S' al valor entero 1230. El carácter S indica un decimal zonificado firmado. Para obtener más información, consulte Caracteres de formato de estilo Teradata para datos numéricos.

```
select text_to_int_alt('123{', '999S');
```

```
text_to_int_alt
----------
      1230
```

En el siguiente ejemplo, se convierte la cadena de expresión de entrada 'USD123' con la frase de formato 'C9(I)' para el valor entero 123. Consulte Caracteres de formato de estilo Teradata para datos numéricos.

```
select text_to_int_alt('USD123', 'C9(I)');
```

```
text_to_int_alt
----------
       123
```

En el siguiente ejemplo, se especifica la columna de una tabla como la expresión de entrada.

```
select text_to_int_alt(a), text_to_int_alt(b) from t_text2int order by 1;
```

```
 text_to_int_alt | text_to_int_alt
-----------------+-----------------
            -123 |            -123
            -123 |            -123
             123 |             123
             123 |             123
```

A continuación, se presenta la definición de la tabla y la instrucción de inserción para este ejemplo.

```
create table t_text2int (a varchar(200), b char(200));
```

```
insert into t_text2int VALUES('123', '123'),('123.123', '123.123'), ('-123', '-123'),
 ('123-', '123-');
```

## TEXT_TO_NUMERIC_ALT

TEXT_TO_NUMERIC_ALT realiza una operación de conversión de estilo Teradata para convertir una cadena de caracteres a un formato de datos numérico.

Sintaxis

```
TEXT_TO_NUMERIC_ALT (expression [, 'format'] [, precision, scale])
```

### Argumentos

expresión

Una expresión que toma uno o más valores CHAR o VARCHAR, como un nombre de columna o un literal. La conversión de valores nulos devuelve valores nulos. Las cadenas en blanco o vacías se convierten a 0.

format

Un literal de cadena que define el formato de la expresión de entrada. Para obtener más información, consulte Caracteres de formato de estilo Teradata para datos numéricos.

precisión

La cantidad de dígitos en el resultado numérico. El valor predeterminado es 38.

scale

La cantidad de dígitos a la derecha de la coma decimal en el resultado numérico. El valor predeterminado es 0.

### Tipo de retorno

TEXT_TO_NUMERIC_ALT devuelve un número DECIMAL.

Amazon Redshift devuelve un error si la conversión a la frase de formato que se especifica no consigue resultados correctos.

Amazon Redshift convierte la cadena de expresión de entrada al tipo numérico con la mayor precisión que especifique para ese tipo en la opción de precisión. Si la longitud del valor numérico supera el valor especificado para la precisión, Amazon Redshift redondea el valor numérico de acuerdo con las siguientes reglas:

- Si la longitud del resultado de conversión excede la longitud especificada en la frase de formato, Amazon Redshift devuelve un error.
- Si el resultado se convierte en un valor numérico, se redondea al valor más cercano. Si la parte de fracción está exactamente a la mitad del resultado de la conversión superior y la inferior, el resultado se redondea al valor par más cercano.

Ejemplos

En el siguiente ejemplo, se convierte la cadena de expresión de entrada '1,5' al valor numérico '2'. Debido a que la instrucción no especifica la escala, la escala adopta el valor predeterminado 0 y el resultado de la conversión no incluye el resultado de una fracción. Como ,5 está a la mitad de 1 y 2, el resultado de la conversión se redondea al valor par de 2.

```
select text_to_numeric_alt('1.5');
```

```
 text_to_numeric_alt
---------------------
                   2
```

En el siguiente ejemplo, se convierte la cadena de expresión de entrada '2,51' al valor numérico 3. Debido a que la instrucción no especifica un valor de escala, la escala adopta el valor predeterminado 0 y el resultado de la conversión no incluye el resultado de una fracción. Como ,51 está más cerca de 3 que de 2, el resultado de la conversión se redondea al valor de 3.

```
select text_to_numeric_alt('2.51');
```

```
 text_to_numeric_alt
---------------------
                   3
```

En el siguiente ejemplo, se convierte la cadena de expresión de entrada 123,52501 con una precisión de 10 y una escala de 2 al valor numérico 123,53.

```
select text_to_numeric_alt('123.52501', 10, 2);
```

```
 text_to_numeric_alt
---------------------
              123.53
```

En el siguiente ejemplo, se convierte la cadena de expresión de entrada "123{" con la frase de formato "999S" al valor numérico 1230. El carácter S indica un decimal zonificado firmado. Para obtener más información, consulte Caracteres de formato de estilo Teradata para datos numéricos.

```
select text_to_numeric_alt('123{', '999S');
```

```
text_to_int_alt
----------
      1230
```

En el siguiente ejemplo, se convierte la cadena de expresión de entrada "USD123" con la frase de formato "C9(I)" al valor numérico 124. Consulte Caracteres de formato de estilo Teradata para datos numéricos.

```
select text_to_numeric_alt('USD123.9', 'C9(I)');
```

```
text_to_numeric_alt
----------
       124
```

En el siguiente ejemplo, se especifica la columna de una tabla como la expresión de entrada.

```
select text_to_numeric_alt(a), text_to_numeric_alt(b) from t_text2numeric order by 1;
```

```
          text_to_numeric_alt           |          text_to_numeric_alt
----------------------------------------+----------------------------------------
 -99999999999999999999999999999999999999 | -99999999999999999999999999999999999999
                                  -12300 |                                  -12300
                                     123 |                                     123
                                     123 |                                     123
  99999999999999999999999999999999999999 |  99999999999999999999999999999999999999
```

A continuación, se presenta la definición de la tabla y la instrucción de inserción para este ejemplo.

```
create table  t_text2numeric (a varchar(200), b char(200));
```

```
insert into  t_text2numeric values
('123', '123'),
('+123.456', '+123.456'),
('-' || repeat('9', 38), '-' || repeat('9', 38)),
(repeat('9', 38) || '+', repeat('9', 38) || '+'),
('-123E2', '-123E2');
```

# TO_CHAR

TO_CHAR convierte una marca temporal o una expresión numérica a un formato de datos de cadena de caracteres.

## Sintaxis

```
TO_CHAR (timestamp_expression | numeric_expression , 'format')
```

## Argumentos

timestamp_expression

Una expresión que da lugar a un valor de tipo TIMESTAMP o TIMESTAMPTZ, o bien, un valor que se pueda convertir de forma implícita en una marca temporal.

numeric_expression

Una expresión que de como resultado un valor de tipo de datos numérico o un valor que se pueda convertir implícitamente en un tipo numérico. Para obtener más información, consulte Tipos numéricos. TO_CHAR inserta un espacio a la izquierda de la cadena numérica.

> **Note**
>
> TO_CHAR no admite valores DECIMAL de 128 bits.

format

El formato para el valor nuevo. Para conocer los formatos válidos, consulte Cadenas de formatos de fecha y hora y Cadenas de formatos numéricos.

## Tipo de retorno

VARCHAR

## Ejemplos

En el ejemplo siguiente, se convierte una marca temporal en un valor con la fecha y la hora en un formato con el nombre del mes relleno con nueve caracteres, el nombre del día de la semana y el número de día del mes.

```
select to_char(timestamp '2009-12-31 23:15:59', 'MONTH-DY-DD-YYYY HH12:MIPM');

to_char
-------------------------
DECEMBER -THU-31-2009 11:15PM
```

En el siguiente ejemplo, se convierte una marca temporal en un valor con el número de día del año.

```
select to_char(timestamp '2009-12-31 23:15:59', 'DDD');

to_char
-------------------------
365
```

En el siguiente ejemplo, se convierte una marca temporal en un número de día de ISO de la semana.

```
select to_char(timestamp '2022-05-16 23:15:59', 'ID');

to_char
-------------------------
1
```

El siguiente ejemplo extrae el nombre del mes de una fecha.

```
select to_char(date '2009-12-31', 'MONTH');

to_char
-------------------------
DECEMBER
```

En el siguiente ejemplo, se convierte cada valor STARTTIME en la tabla EVENT a una cadena que consta de horas, minutos y segundos.

```
select to_char(starttime, 'HH12:MI:SS')
from event where eventid between 1 and 5
order by eventid;

to_char
----------
02:30:00
```

```
08:00:00
02:30:00
02:30:00
07:00:00
```

En el siguiente ejemplo, se convierte un valor completo de marca temporal a un formato diferente.

```
select starttime, to_char(starttime, 'MON-DD-YYYY HH12:MIPM')
from event where eventid=1;

      starttime      |       to_char
---------------------+---------------------
 2008-01-25 14:30:00 | JAN-25-2008 02:30PM
```

En el siguiente ejemplo, se convierte un literal de marca temporal a una cadena de caracteres.

```
select to_char(timestamp '2009-12-31 23:15:59','HH24:MI:SS');

to_char
----------
23:15:59
```

En el siguiente ejemplo, se convierte un número decimal en una cadena de caracteres.

```
select to_char(125.8, '999.99');

to_char
---------
125.80
```

En el siguiente ejemplo, se convierte un número decimal en una cadena de caracteres.

```
select to_char(125.8, '999D99');

to_char
---------
125.80
```

En el siguiente ejemplo, se convierte un número en una cadena de caracteres con un cero a la izquierda.

```
select to_char(125.8, '0999D99');

to_char
---------
0125.80
```

En el siguiente ejemplo se convierte un número a una cadena de caracteres con el signo negativo al final.

```
select to_char(-125.8, '999D99S');

to_char
---------
125.80-
```

En el siguiente ejemplo se convierte un número a una cadena de caracteres con el signo positivo o negativo en la posición especificada.

```
select to_char(125.8, '999D99SG');

to_char
---------
125.80+
```

En el siguiente ejemplo se convierte un número a una cadena de caracteres con el signo positivo en la posición especificada.

```
select to_char(125.8, 'PL999D99');

to_char
---------
+ 125.80
```

En el siguiente ejemplo se convierte un número a una cadena de caracteres con el símbolo de moneda.

```
select to_char(-125.88, '$S999D99');

to_char
---------
```

```
$-125.88
```

En el siguiente ejemplo se convierte un número a una cadena de caracteres con el símbolo de moneda en la posición especificada.

```
select to_char(-125.88, 'S999D99L');

to_char
---------
-125.88$
```

En el siguiente ejemplo, se convierte un número a una cadena de caracteres utilizando un separador de millar (coma).

```
select to_char(1125.8, '9,999.99');

to_char
---------
1,125.80
```

En el siguiente ejemplo, se convierte un número a una cadena de caracteres con corchetes angulares para números negativos.

```
select to_char(-125.88, '$999D99PR');

to_char
---------
$<125.88>
```

En el siguiente ejemplo se convierte un número a una cadena de números romanos.

```
select to_char(125, 'RN');

to_char
---------
   CXXV
```

En el siguiente ejemplo, se convierte una fecha a un código de siglo.

```
select to_char(date '2020-12-31', 'CC');
```

```
to_char
---------
21
```

En el ejemplo siguiente se muestra el día de la semana.

```
SELECT to_char(current_timestamp, 'FMDay, FMDD HH12:MI:SS');

to_char
-----------------------
Wednesday, 31 09:34:26
```

En el ejemplo siguiente se muestra el sufijo de número ordinal de un número.

```
SELECT to_char(482, '999th');

to_char
-----------------------
 482nd
```

En el siguiente ejemplo, se resta la comisión del precio pagado en la tabla de ventas. La diferencia, luego, se redondea hacia arriba y se convierte en un número romano, que se muestra en la columna to_char:

```
select salesid, pricepaid, commission, (pricepaid - commission)
as difference, to_char(pricepaid - commission, 'rn') from sales
group by sales.pricepaid, sales.commission, salesid
order by salesid limit 10;

 salesid | pricepaid | commission | difference |     to_char
---------+-----------+------------+------------+-----------------
       1 |    728.00 |     109.20 |     618.80 |           dcxix
       2 |     76.00 |      11.40 |      64.60 |             lxv
       3 |    350.00 |      52.50 |     297.50 |        ccxcviii
       4 |    175.00 |      26.25 |     148.75 |           cxlix
       5 |    154.00 |      23.10 |     130.90 |           cxxxi
       6 |    394.00 |      59.10 |     334.90 |         cccxxxv
       7 |    788.00 |     118.20 |     669.80 |           dclxx
       8 |    197.00 |      29.55 |     167.45 |          clxvii
       9 |    591.00 |      88.65 |     502.35 |             dii
      10 |     65.00 |       9.75 |      55.25 |              lv
```

En el siguiente ejemplo, se agrega el símbolo de la moneda a los valores de diferencia que se muestran en la columna to_char:

```
select salesid, pricepaid, commission, (pricepaid - commission)
as difference, to_char(pricepaid - commission, 'l99999D99') from sales
group by sales.pricepaid, sales.commission, salesid
order by salesid limit 10;

salesid | pricepaid | commission | difference |  to_char
--------+-----------+------------+------------+------------
      1 |    728.00 |     109.20 |     618.80 | $   618.80
      2 |     76.00 |      11.40 |      64.60 | $    64.60
      3 |    350.00 |      52.50 |     297.50 | $   297.50
      4 |    175.00 |      26.25 |     148.75 | $   148.75
      5 |    154.00 |      23.10 |     130.90 | $   130.90
      6 |    394.00 |      59.10 |     334.90 | $   334.90
      7 |    788.00 |     118.20 |     669.80 | $   669.80
      8 |    197.00 |      29.55 |     167.45 | $   167.45
      9 |    591.00 |      88.65 |     502.35 | $   502.35
     10 |     65.00 |       9.75 |      55.25 | $    55.25
```

En el siguiente ejemplo, se indica el siglo en el que se realizó la venta.

```
select salesid, saletime, to_char(saletime, 'cc') from sales
order by salesid limit 10;

 salesid |      saletime       | to_char
---------+---------------------+---------
       1 | 2008-02-18 02:36:48 | 21
       2 | 2008-06-06 05:00:16 | 21
       3 | 2008-06-06 08:26:17 | 21
       4 | 2008-06-09 08:38:52 | 21
       5 | 2008-08-31 09:17:02 | 21
       6 | 2008-07-16 11:59:24 | 21
       7 | 2008-06-26 12:56:06 | 21
       8 | 2008-07-10 02:12:36 | 21
       9 | 2008-07-22 02:23:17 | 21
      10 | 2008-08-06 02:51:55 | 21
```

En el siguiente ejemplo, se convierte cada valor STARTTIME en la tabla EVENT en una cadena que consta de horas, minutos, segundos y zona horaria.

```
select to_char(starttime, 'HH12:MI:SS TZ')
from event where eventid between 1 and 5
order by eventid;

to_char
----------
02:30:00 UTC
08:00:00 UTC
02:30:00 UTC
02:30:00 UTC
07:00:00 UTC
```

En el siguiente ejemplo, se muestra el formato para segundos, milisegundos y microsegundos.

```
select sysdate,
to_char(sysdate, 'HH24:MI:SS') as seconds,
to_char(sysdate, 'HH24:MI:SS.MS') as milliseconds,
to_char(sysdate, 'HH24:MI:SS:US') as microseconds;

timestamp           | seconds  | milliseconds | microseconds
--------------------+----------+--------------+----------------
2015-04-10 18:45:09 | 18:45:09 | 18:45:09.325 | 18:45:09:325143
```

## Función TO_DATE

TO_DATE convierte una fecha que se representa con una cadena de caracteres en un tipo de datos DATE.

> **Note**
>
> TO_DATE no admite cadenas de formato con Q (número de trimestre).

### Sintaxis

```
TO_DATE(string, format)
```

```
TO_DATE(string, format, is_strict)
```

Argumentos

string

La cadena que se convertirá.

format

Un literal de cadena que define el formato de la entrada cadena, en términos de sus partes de fecha. Para obtener una lista de los formatos válidos para día, mes y año, consulte Cadenas de formatos de fecha y hora.

is_strict

Un valor booleano opcional que especifica si se devuelve un error si un valor de fecha de entrada se encuentra fuera de rango. Cuando is_strict se configura como TRUE, se devuelve un error si hay un valor fuera de rango. Si is_strict se configura como FALSE, que es el valor predeterminado, se aceptan valores de desbordamiento.

Tipo de retorno

TO_DATE devuelve un valor DATE, en función del valor de format.

Si la conversión a formato produce un error, se devuelve un error.

Ejemplos

La siguiente instrucción SQL convierte la fecha `02 Oct 2001` a un tipo de datos de fecha.

```
select to_date('02 Oct 2001', 'DD Mon YYYY');

to_date
------------
2001-10-02
(1 row)
```

La siguiente instrucción SQL convierte la cadena `20010631` en una fecha.

```
select to_date('20010631', 'YYYYMMDD', FALSE);
```

El resultado es 1.° de julio de 2001, ya que solo hay 30 días en junio.

```
to_date
------------
2001-07-01
```

La siguiente instrucción SQL convierte la cadena `20010631` en una fecha:

```
to_date('20010631', 'YYYYMMDD', TRUE);
```

El resultado es de error, ya que solo hay 30 días en junio.

```
ERROR:  date/time field date value out of range: 2001-6-31
```

## TO_NUMBER

TO_NUMBER convierte una cadena en un valor numérico (decimal).

> **Note**
>
> Le recomendamos que utilice `FM` en la cadena de formato para suprimir el relleno de ceros y espacios en blanco. Para obtener una lista de formatos válidos, consulte Cadenas de formatos numéricos.

### Sintaxis

```
to_number(string, format)
```

### Argumentos

string

Cadena que se convertirá. El formato debe ser un valor literal.

format

El segundo argumento es una cadena de formato que indica cómo se debe analizar la cadena original para crear el valor numérico. Por ejemplo, el formato `'FM99D999'` especifica que la cadena que se convertirá consta de cinco dígitos con el punto decimal en la tercera posición. Por

ejemplo, `to_number('12.345','FM99D999')` devuelve `12.345` como un valor numérico. Para obtener una lista de formatos válidos, consulte Cadenas de formatos numéricos.

Tipo de retorno

TO_NUMBER devuelve un número DECIMAL.

Si la conversión a formato produce un error, se devuelve un error.

Ejemplos

En el siguiente ejemplo, se convierte la cadena `12,454.8-` a un número:

```
select to_number('12,454.8-', 'FM99G999D9S');

to_number
-----------
-12454.8
```

En el siguiente ejemplo, se convierte la cadena `$ 12,454.88` a un número:

```
select to_number('$ 12,454.88', 'FML99G999D99');

to_number
-----------
12454.88
```

En el siguiente ejemplo, se convierte la cadena `$ 2,012,454.88` a un número:

```
select to_number('$ 2,012,454.88', 'FML9,999,999.99');

to_number
-----------
2012454.88
```

## Función TRY_CAST

En comparación con la función CAST, TRY_CAST intenta primero convertir la expresión al tipo especificado. Si la conversión falla debido a errores de conversión, la operación devuelve null. Si una conversión no está permitida explícitamente, la operación devuelve un error. Puede encontrar la lista

de posibles conversiones en las notas de uso que figuran a continuación. Por ejemplo, no se permite intentar convertir un booleano en una marca temporal.

Sintaxis

```
TRY_CAST(expression AS type)
```

Argumentos

expresión

Una expresión que toma el valor de uno o más valores, como un nombre de columna o un literal. La conversión de valores nulos devuelve valores nulos. La expresión no puede tener cadenas en blanco ni vacías.

type

Uno de los tipos de datos admitidos. Para obtener una lista completa de los tipos de datos, consulte Tipos de datos. Para obtener la lista de pares de tipos de datos de origen y tipos de datos de destino admitidos, consulte las notas de uso que figuran a continuación.

Tipo de retorno

TRY_CAST devuelve un valor del tipo de datos especificado por el argumento type. Si la conversión falla, la operación devuelve null.

Notas de uso

A continuación, se muestra la lista de pares de tipos de datos de origen y tipos de datos de destino que Amazon Redshift admite para TRY_CAST.

BOOL

SMALLINT, INT, BIGINT, SUPER

SMALLINT

BOOL, INT, BIGINT, DECIMAL, REAL, FLOAT, BPCHAR, TEXT, VARCHAR, SUPER

INT

BOOL, SMALLINT, BIGINT, DECIMAL, REAL, FLOAT, BPCHAR, TEXT, VARCHAR, SUPER

BIGINT

BOOL, SMALLINT, INT, DECIMAL, REAL, FLOAT, BPCHAR, TEXT, VARCHAR, SUPER

DECIMAL

SMALLINT, INT, BIGINT, REAL, FLOAT, BPCHAR, TEXT, VARCHAR, SUPER

REAL

SMALLINT, INT, BIGINT, DECIMAL, FLOAT, BPCHAR, TEXT, VARCHAR, SUPER

FLOAT

SMALLINT, INT, BIGINT, DECIMAL, REAL, BPCHAR, TEXT, VARCHAR, SUPER

BPCHAR

SMALLINT, INT, BIGINT, DECIMAL, REAL, FLOAT, TEXT, VARCHAR, TIMESTAMP, TIMESTAMPTZ, DATE, TIME, TIMETZ, SUPER

TEXT

SMALLINT, INT, BIGINT, DECIMAL, REAL, FLOAT, BPCHAR, VARCHAR, TIMESTAMP, TIMESTAMPTZ, DATE, TIME, TIMETZ, SUPER

VARCHAR

SMALLINT, INT, BIGINT, DECIMAL, REAL, FLOAT, BPCHAR, TEXT, TIMESTAMP, TIMESTAMPTZ, DATE, TIME, TIMETZ, SUPER

TIMESTAMP

BPCHAR, TEXT, VARCHAR, TIMESTAMPTZ, DATE, TIME

TIMESTAMPTZ

BPCHAR, TEXT, VARCHAR, TIMESTAMP, DATE, TIME, TIMETZ

DATE

BPCHAR, TEXT, VARCHAR, TIMESTAMP, TIMESTAMPTZ

TIME

BPCHAR, TEXT, VARCHAR

TIMETZ

BPCHAR, TEXT, VARCHAR

SUPER

SUPER se puede convertir a cualquier otro tipo de datos, con la excepción de DATE, TIME, TIMETZ, TIMESTAMP y TIMESTAMPTZ.

Ejemplos

El siguiente ejemplo convierte un STRING en un INTEGER.

```
SELECT TRY_CAST('123' AS INT);

int
----
123
```

El ejemplo siguiente devuelve null. Se permite convertir un STRING en un INTEGER, por lo que TRY_CAST no devuelve un error, pero “foo” no es un entero, por lo que la función devuelve null.

```
SELECT TRY_CAST('foo' AS INT)
```

El siguiente ejemplo devuelve un error, porque no se permite convertir un BOOLEAN en un TIMESTAMP.

```
SELECT TRY_CAST(true as timestamp);
```

Debido a que TRY_CAST devuelve un valor nulo en lugar de devolver inmediatamente un error si la conversión falla, puede utilizar TRY_CAST para filtrar datos no válidos. Considere el siguiente ejemplo, en el que se filtra una fila no válida debido a un error de conversión en la columna de edad de Akua Mansa.

```
CREATE TABLE IF NOT EXISTS student_data(
name VARCHAR(100) NOT NULL,
age VARCHAR(3) NOT NULL,
enrollment_date DATE NOT NULL);

INSERT INTO student_data (name, age, enrollment_date)
VALUES
('Alejandro Rosalez', '10', '01/01/2000'),
('Akua Mansa', 'Ten', '01/01/2000');
```

```
SELECT * FROM student_data WHERE TRY_CAST(age AS INT) IS NOT NULL;

--Akua is not returned.
 name              | age | enrollment_date
-------------------+-----+-----------------
 Alejandro Rosalez | 10  | 01/01/2000
```

## Cadenas de formatos de fecha y hora

Puede encontrar una referencia para las siguientes cadenas de formato de fecha y hora.

Las siguientes cadenas de formato se aplican a funciones como TO_CHAR. Estas cadenas pueden tener separadores de fecha y hora (como "-", "/" o ":·) y las siguientes "partes de fecha" y "partes de hora".

| Partes de fecha o de hora | Significado |
|---|---|
| BC o B.C., AD o A.D., b.c. o bc, ad o a.d. | Indicadores de era en mayúsculas o minúsculas |
| CC | Número de siglo en dos dígitos |
| YYYY, YYY, YY, Y | Número de año en 4 dígitos, 3 dígitos, 2 dígitos, 1 dígito |
| Y,YYY | Número de año en 4 dígitos con coma |
| IYYY, IYY, IY, I | Número de año según la Organización Internacional de Normalización (ISO), en 4 dígitos, 3 dígitos, 2 dígitos y 1 dígito |
| Q | Número de trimestre (1 a 4) |
| MONTH, Month, month | Nombre del mes (mayúsculas, mayúsculas y minúsculas, minúsculas, con un espacio en blanco de 9 caracteres) |
| MON, Mon, mon | Nombre del mes abreviado (mayúsculas, mayúsculas y minúsculas, minúsculas, con un espacio en blanco de 3 caracteres) |

| Partes de fecha o de hora | Significado |
| --- | --- |
| MM | Número de mes (01-12) |
| RM, rm | Número de mes en números romanos (I a XII, donde I es enero, ya sea en mayúsculas o minúsculas) |
| W | Semana del mes (1 a 5; la primera semana comienza el primer día del mes). |
| WW | Número de semana del año (1 a 53; la primera semana comienza el primer día del año). |
| IW | Número de semana del año según la ISO (el primer jueves del nuevo año está en la semana 1.) |
| DAY, Day, day | Nombre del día (mayúsculas, mayúsculas y minúsculas, minúsculas, con un espacio en blanco de 9 caracteres) |
| DY, Dy, dy | Nombre del día abreviado (mayúsculas, mayúsculas y minúsculas, minúsculas, con un espacio en blanco de 3 caracteres) |
| DDD | Día del año (001 a 366) |
| IDDD | Día del año según la numeración de semanas de ISO 8601 (001-371; el primer día del año es el lunes de la primera semana ISO) |
| DD | Día del mes como un número (01 a 31) |

| Partes de fecha o de hora | Significado |
| --- | --- |
| D | Día de la semana (1 a 7; el domingo es el día 1)<br><br>**Note**<br>La parte de fecha D se comporta de manera diferente a la parte de fecha día (DOW) que se usa para las funciones de fecha y hora DATE_PART y EXTRACT. DOW se basa en los números enteros de 0 a 6, con el domingo como 0. Para obtener más información, consulte Partes de fecha para funciones de fecha o marca temporal. |
| ID | Día de la semana de ISO 8601, de lunes (1) a domingo (7) |
| J | Día juliano (días desde el 1.º de enero de 4712 AC) |
| HH24 | Hora (formato de 24 horas, de 00 a 23) |
| HH o HH12 | Hora (formato de 12 horas, de 01 a 12) |
| MI | Minutos (00 a 59) |
| SS | Segundos (00 a 59) |
| MS | Milisegundos (0,000) |
| US | Microsegundos (0,000000) |
| AM o PM, A.M. o P.M., a.m. o p.m., am o pm | Indicadores meridianos en mayúsculas o minúsculas (para formato de 12 horas) |

| Partes de fecha o de hora | Significado |
| --- | --- |
| TZ, tz | Abreviación de la zona horaria en mayúsculas o minúsculas; válido solo para TIMESTAMPTZ |
| OF | Desplazamiento del huso UTC; válido solo para TIMESTAMPTZ |

> **Note**
>
> Tiene que incluir los separadores de datetime (como '-', '/' o ':') entre comillas simples, pero los "dateparts" y los "timeparts" enumerados en la tabla anterior tienen que estar entre comillas dobles.

Ejemplos

Para ver ejemplos de cómo formatear fechas como cadenas, consulte TO_CHAR.

## Cadenas de formatos numéricos

A continuación, puede encontrar una referencia para cadenas de formato numérico.

Las siguientes cadenas de formato se aplican a funciones como TO_NUMBER y TO_CHAR.

- Para ver ejemplos de cómo formatear cadenas como números, consulte TO_NUMBER.
- Para ver ejemplos de cómo formatear números como cadenas, consulte TO_CHAR.

| Formato | Descripción |
| --- | --- |
| 9 | Valor numérico con la cantidad especificada de dígitos. |
| 0 | Valor numérico con ceros a la izquierda. |
| . (period), D | Punto decimal. |
| , (coma) | Separador de miles. |

| Formato | Descripción |
| --- | --- |
| CC | Código de siglo. Por ejemplo, el siglo XXI comenzó el 01/01/2001 (compatible solo con TO_CHAR). |
| FM | Modo de relleno. Suprime espacios de relleno y ceros. |
| PR | Valor negativo entre paréntesis. |
| S | Signo anclado a un número. |
| L | El símbolo de la moneda en la posición especificada. |
| G | Separador de grupo. |
| MI | Signo menos en la posición especificada para números menores que 0. |
| PL | Signo más en la posición especificada para números mayores que 0. |
| SG | Signo más o menos en la posición especific ada. |
| RN | Número romano entre 1 y 3999 (compatible solo con TO_CHAR). |
| TH o th | Sufijo de número ordinal. No convierte fracciones ni valores menores que cero. |

## Caracteres de formato de estilo Teradata para datos numéricos

A continuación, puede descubrir cómo las funciones TEXT_TO_INT_ALT y TEXT_TO_NUMERIC_ALT interpretan los caracteres de la cadena de expresión de entrada. También encontrará una lista con los caracteres que puede especificar en la frase de formato. Además,

encontrará una descripción de las diferencias entre el formato de estilo Teradata y Amazon Redshift para la opción de formato.

| Formato | Descripción |
| --- | --- |
| G | No se admite como separador de grupo en la cadena de expresión de entrada. No puede especificar este carácter en la frase de formato. |
| D | Símbolo de punto base. Puede especificar este carácter en la frase de formato. Este carácter equivale al . (punto).<br><br>El símbolo de punto base no puede aparecer en una frase de formato que contenga cualquiera de los siguientes caracteres:<br><br>• . (punto)<br>• S (“s” mayúscula)<br>• V (“v” mayúscula) |
| / , : % | Caracteres de inserción: barra diagonal (/), coma (,), dos puntos (:) y signo de porcentaje (%).<br><br>No puede incluir estos caracteres en la frase de formato.<br><br>Amazon Redshift no tiene en cuenta estos caracteres en la cadena de expresión de entrada. |
| . | Un punto como un carácter de punto base, es decir, una coma decimal.<br><br>Este carácter no puede aparecer en una frase de formato que contenga cualquiera de los siguientes caracteres: |

| Formato | Descripción |
| --- | --- |
| | • D (“d” mayúscula)<br>• S (“s” mayúscula)<br>• V (“v” mayúscula) |
| B | No puede incluir el carácter de espacio en blanco (B) en la frase de formato. En la cadena de expresión de entrada, los espacios anteriores y posteriores no se tienen en cuenta, y los espacios entre dígitos no están permitidos. |
| + - | No puede incluir el signo más (+) o el signo menos (-) en la frase de formato. No obstante, el signo más (+) y el signo menos (-) se analizan de forma implícita como parte del valor numérico si aparecen en la cadena de expresión de entrada. |
| V | Indicador de la posición de la coma decimal.<br><br>Este carácter no puede aparecer en una frase de formato que contenga cualquiera de los siguientes caracteres:<br><br>• D (“d” mayúscula)<br>• . (punto) |
| Z | Dígito decimal con los ceros suprimidos. Amazon Redshift quita los ceros a la izquierda . El carácter Z no puede seguir a un carácter 9. El carácter Z debe estar a la izquierda del carácter de punto base si la parte de la fracción contiene el carácter 9. |
| 9 | Dígito decimal. |

| Formato | Descripción |
| --- | --- |
| CHAR(n) | Para este formato, puede especificar lo siguiente:<br>• CHAR consta de los caracteres Z o 9. Amazon Redshift no admite un + (más) o un - (menos) en el valor CHAR.<br>• n es una constante entera, I o F. Para I, es el número de caracteres necesarios para mostrar la parte entera de los datos numéricos o enteros. Para F, es el número de caracteres necesarios para mostrar la parte de fracción de los datos numéricos. |
| - | Carácter de guion (-).<br>No puede incluir este carácter en la frase de formato.<br>Amazon Redshift no tiene en cuenta este carácter en la cadena de expresión de entrada. |

| Formato | Descripción |
| --- | --- |
| S | Decimal zonificado firmado. El carácter S debe ubicarse después del último dígito decimal de la frase de formato. El último carácter de la cadena de expresión de entrada y la conversió n numérica correspondiente se mencionan en Caracteres de formato de datos para un decimal zonificado firmado y formato de datos numéricos de estilo Teradata .<br><br>El carácter S no puede aparecer en una frase de formato que contenga cualquiera de los siguientes caracteres:<br><br>• + (signo más)<br>• . (punto)<br>• D (“d” mayúscula)<br>• Z (“z” mayúscula)<br>• F (“f” mayúscula)<br>• E (“e” mayúscula) |
| E | Notación exponencial. La cadena de expresión de entrada puede incluir el carácter de exponente. No se puede especificar E como un carácter de exponente en la frase de formato. |
| FN9 | No es compatible con Amazon Redshift. |
| FNE | No es compatible con Amazon Redshift. |

| Formato | Descripción |
| --- | --- |
| $, USD, dólares estadounidenses | Signo de dólar ($), código ISO de la moneda (USD) y nombre de la moneda dólares estadounidenses.<br><br>El código ISO de la moneda, USD, y el nombre de la moneda, dólares estadounidenses, distinguen entre mayúsculas y minúscula s. Amazon Redshift solo admite la moneda USD. La entrada de expresión de entrada puede incluir espacios entre el símbolo de la moneda USD y el valor numérico; por ejemplo, “$ 123E2” o “123E2 $”. |
| L | Símbolo de la moneda. Este carácter de símbolo de moneda solo puede aparecer una vez en la frase de formato. No se pueden especificar caracteres de símbolo de moneda repetidos. |
| C | Código de la moneda ISO. Este carácter de símbolo de moneda solo puede aparecer una vez en la frase de formato. No se pueden especificar caracteres de símbolo de moneda repetidos. |
| N | Nombre completo de la moneda. Este carácter de símbolo de moneda solo puede aparecer una vez en la frase de formato. No se pueden especificar caracteres de símbolo de moneda repetidos. |
| O | Símbolo de moneda doble. No puede especific ar este carácter en la frase de formato. |
| U | Código de moneda ISO doble. No puede especificar este carácter en la frase de formato. |

| Formato | Descripción |
| --- | --- |
| A | Nombre completo doble de la moneda. No puede especificar este carácter en la frase de formato. |

Caracteres de formato de datos para un decimal zonificado firmado y formato de datos numéricos de estilo Teradata

Puede utilizar los siguientes caracteres en la frase de formato de las funciones TEXT_TO_INT_ALT y TEXT_TO_NUMERIC_ALT para un valor decimal, zonificado y firmado.

| Último carácter de la cadena de entrada | Conversión numérica |
| --- | --- |
| { o 0 | n … 0 |
| A o 1 | n … 1 |
| B o 2 | n … 2 |
| C o 3 | n … 3 |
| D o 4 | n … 4 |
| E o 5 | n … 5 |
| F o 6 | n … 6 |
| G o 7 | n … 7 |
| H u 8 | n … 8 |
| I o 9 | n … 9 |
| } | -n … 0 |
| J | -n … 1 |
| K | -n … 2 |

| Último carácter de la cadena de entrada | Conversión numérica |
|---|---|
| L | -n … 3 |
| M | -n … 4 |
| N | -n … 5 |
| O | -n … 6 |
| P | -n … 7 |
| Q | -n … 8 |
| R | -n … 9 |

## Funciones de fecha y hora

En esta sección, puede encontrar información acerca de las funciones escalares de fecha y hora compatibles con Amazon Redshift.

Temas


- Resumen de las funciones de fecha y hora
- Funciones de fecha y hora en transacciones
- Funciones específicas del nodo principal obsoletas
- + Operador (concatenación)
- Función ADD_MONTHS
- Función AT TIME ZONE
- Función CONVERT_TIMEZONE
- Función CURRENT_DATE
- Función DATE_CMP
- Función DATE_CMP_TIMESTAMP
- Función DATE_CMP_TIMESTAMPTZ
- Función DATEADD
- Función DATEDIFF

- Función DATE_PART
- Función DATE_PART_YEAR
- Función DATE_TRUNC
- Función EXTRACT
- Función GETDATE
- Función INTERVAL_CMP
- Función LAST_DAY
- Función MONTHS_BETWEEN
- Función NEXT_DAY
- Función SYSDATE
- Función TIMEOFDAY
- Función TIMESTAMP_CMP
- Función TIMESTAMP_CMP_DATE
- Función TIMESTAMP_CMP_TIMESTAMPTZ
- Función TIMESTAMPTZ_CMP
- Función TIMESTAMPTZ_CMP_DATE
- Función TIMESTAMPTZ_CMP_TIMESTAMP
- Función TIMEZONE
- Función TO_TIMESTAMP
- Función TRUNC
- Partes de fecha para funciones de fecha o marca temporal

## Resumen de las funciones de fecha y hora

| Función | Sintaxis | Devuelve |
| --- | --- | --- |
| + Operador (concatenación)<br><br>Concatena una fecha a una hora a cada lado del símbolo + y devuelve TIMESTAMP o TIMESTAMPTZ. | date + time | `TIMESTAMP` o `TIMESTAMPTZ` |

| Función | Sintaxis | Devuelve |
| --- | --- | --- |
| ADD_MONTHS<br><br>Agrega la cantidad de meses especificada a una fecha o marca temporal. | ADD_MONTHS ({date\|timestamp}, integer) | TIMESTAMP |
| AT TIME ZONE<br><br>Especifica la zona horaria que debe usarse con una expresión TIMESTAMP o TIMESTAMPTZ. | AT TIME ZONE 'timezone' | TIMESTAMP o TIMESTAMPTZ |
| CONVERT_TIMEZONE<br><br>Convierte una marca temporal de una zona horaria a otra. | CONVERT_TIMEZONE (['timezone',] 'timezone', timestamp) | TIMESTAMP |
| CURRENT_DATE<br><br>Devuelve una fecha en la zona horaria de la sesión actual (que es UTC de manera predeterminada) para el comienzo de la transacción actual. | CURRENT_DATE | DATE |
| DATE_CMP<br><br>Compara dos fechas y devuelve 0 si son idénticas, 1 si date1 es posterior y -1 si date2 es posterior. | DATE_CMP (date1, date2) | INTEGER |
| DATE_CMP_TIMESTAMP<br><br>Compara una fecha y una hora, y devuelve 0 si los valores son idénticos, 1 si date es posterior y -1 si timestamp es posterior. | DATE_CMP_TIMESTAMP (date, timestamp) | INTEGER |

| Función | Sintaxis | Devuelve |
| --- | --- | --- |
| DATE_CMP_TIMESTAMPTZ<br><br>Compara una fecha y una marca temporal con una zona horaria, y devuelve `0` si los valores son idénticos, `1` si date es posterior y `-1` si timestamptz es posterior. | DATE_CMP_TIMESTAMPTZ (date, timestamptz) | `INTEGER` |
| DATE_PART_YEAR<br><br>Extrae el año de una fecha. | DATE_PART_YEAR (date) | `INTEGER` |
| DATEADD<br><br>Aumenta una fecha o una hora según un intervalo especificado. | DATEADD (datepart, interval, {date\|time\|timetz\|timestamp}) | `TIMESTAMP` o `TIME` o `TIMETZ` |
| DATEDIFF<br><br>Devuelve la diferencia entre dos fechas u horas para una parte de fecha dada, como un día o mes. | DATEDIFF (datepart, {date\|time\|timetz\|timestamp}, {date\|time\|timetz\|timestamp}) | `BIGINT` |
| DATE_PART<br><br>Extrae el valor de la parte de una fecha a partir de una fecha u hora. | DATE_PART (datepart, {date\|timestamp}) | `DOUBLE` |
| DATE_TRUNC<br><br>Trunca una marca temporal en función de una parte de fecha. | DATE_TRUNC ('datepart', timestamp) | `TIMESTAMP` |
| EXTRACT<br><br>Extrae una parte de una fecha o una hora a partir de timestamp, timestamptz, time o timetz. | EXTRACT (parte de fecha FROM origen) | `INTEGER or DOUBLE` |

| Función | Sintaxis | Devuelve |
| --- | --- | --- |
| GETDATE<br><br>Devuelve la fecha y hora actual en la zona horaria de la sesión actual (que es UTC de manera predeterminada). Los paréntesis son obligatorios. | GETDATE() | TIMESTAMP |
| INTERVAL_CMP<br><br>Compara dos intervalos y devuelve 0 si son iguales, 1 si interval1 es posterior y -1 si interval2 es posterior. | INTERVAL_CMP (interval1, interval2) | INTEGER |
| LAST_DAY<br><br>Devuelve la fecha del último día del mes que contiene el valor de date (fecha). | LAST_DAY(date) | DATE |
| MONTHS_BETWEEN<br><br>Devuelve la cantidad de meses que hay entre dos fechas. | MONTHS_BETWEEN (date, date) | FLOAT8 |
| NEXT_DAY<br><br>Devuelve la fecha de la primera instancia de day (día) posterior a date (fecha). | NEXT_DAY (date, day) | DATE |
| SYSDATE<br><br>Devuelve la fecha y hora según la zona horaria UTC para el comienzo de la transacción actual. | SYSDATE | TIMESTAMP |

| Función | Sintaxis | Devuelve |
| --- | --- | --- |
| TIMEOFDAY<br><br>Devuelve el día, la fecha y la hora actuales en la zona horaria de la sesión actual (que es UTC de manera predeterminada) como un valor de cadena. | TIMEOFDAY() | VARCHAR |
| TIMESTAMP_CMP<br><br>Compara dos marcas temporales y devuelve 0 si son iguales, 1 si timestamp1 (marca_temporal1) es posterior y -1 si timestamp2 (marca_temporal2) es posterior. | TIMESTAMP_CMP (timestamp1, timestamp2) | INTEGER |
| TIMESTAMP_CMP_DATE<br><br>Compara una marca temporal con una fecha y devuelve 0 si los valores son idénticos, 1 si timestamp es posterior y -1 si date es posterior. | TIMESTAMP_CMP_DATE (timestamp, date) | INTEGER |
| TIMESTAMP_CMP_TIMESTAMPTZ<br><br>Compara una marca temporal y una marca temporal con zona horaria, y devuelve 0 si los valores son iguales, 1 si timestamp es posterior y -1 si timestamptz es posterior. | TIMESTAMP_CMP_TIMESTAMPTZ (timestamp, timestamptz) | INTEGER |
| TIMESTAMPTZ_CMP<br><br>Compara dos marcas temporales con valores de zona horaria y devuelve 0 si los valores son iguales, 1 si timestamptz1 es posterior y -1 si timestamptz2 es posterior. | TIMESTAMPTZ_CMP (timestamptz1, timestamptz2) | INTEGER |

| Función | Sintaxis | Devuelve |
| --- | --- | --- |
| TIMESTAMPTZ_CMP_DATE<br><br>Compara el valor de una marca temporal con zona horaria y una fecha, y devuelve `0` si los valores son iguales, 1 si timestamptz es posterior y `-1` si date es posterior. | TIMESTAMPTZ_CMP_DATE (timestamptz, date) | `INTEGER` |
| TIMESTAMPTZ_CMP_TIMESTAMP<br><br>Compara una marca temporal con zona horaria y una marca temporal, y devuelve `0` si los valores son iguales, 1 si timestamptz es posterior y `-1` si timestamp es posterior. | TIMESTAMPTZ_CMP_TIMESTAMP (timestamptz, timestamp) | `INTEGER` |
| TIMEZONE<br><br>Devuelve una marca temporal para el valor de marca temporal y zona horaria especificados. | TIMEZONE ('timezone' { timestamp \| timestamptz ) | `TIMESTAMP` o `TIMESTAMPTZ` |
| TO_TIMESTAMP<br><br>Devuelve una marca temporal con zona horaria para el formato de marca temporal y zona horaria especificado. | TO_TIMESTAMP ('timestamp', 'format') | `TIMESTAMPTZ` |
| TRUNC<br><br>Trunca una marca temporal y devuelve una fecha. | TRUNC(timestamp) | `DATE` |

> **Note**
>
> Los segundos de salto no se consideran en los cálculos de tiempo transcurrido.

## Funciones de fecha y hora en transacciones

Cuando se ejecutan las siguientes funciones dentro de un bloque de transacción (BEGIN ... END), la función devuelve la fecha u hora de inicio de la transacción actual, no de la instrucción actual.

- SYSDATE
- TIMESTAMP
- CURRENT_DATE

Las siguientes funciones siempre devuelven la fecha u hora de comienzo de la instrucción actual, incluso cuando se encuentran dentro de un bloque de transacción.

- GETDATE
- TIMEOFDAY

## Funciones específicas del nodo principal obsoletas

Las siguientes funciones de fecha quedan obsoletas porque se ejecutan únicamente en el nodo principal. Para obtener más información, consulte Funciones específicas del nodo principal.

- AGE. En su lugar, use Función DATEDIFF.
- CURRENT_TIME. En su lugar, use Función GETDATE o SYSDATE.
- CURRENT_TIMESTAMP. En su lugar, use Función GETDATE o SYSDATE.
- LOCALTIME. En su lugar, use Función GETDATE o SYSDATE.
- LOCALTIMESTAMP. En su lugar, use Función GETDATE o SYSDATE.
- ISFINITE
- NOW. En su lugar, use Función GETDATE o SYSDATE. Si utiliza la función NOW en una vista materializada, se establece en la marca de tiempo de la creación de la vista materializada y no en la marca de tiempo actual.

## + Operador (concatenación)

Concatena una DATE a una TIME o TIMETZ a ambos lados del símbolo + y devuelve una TIMESTAMP o TIMESTAMPTZ.

### Sintaxis

```
date + {time | timetz}
```

El orden de los argumentos se puede invertir. Por ejemplo, time + date.

### Argumentos

date

Una columna de tipo de datos DATE o una expresión que toma implícitamente el valor de un tipo DATE.

time

Una columna de tipo de datos TIME o una expresión que toma implícitamente el valor de un tipo TIME.

timetz

Una columna de tipo de datos TIMETZ o una expresión que toma implícitamente el valor de un tipo TIMETZ.

### Tipo de retorno

TIMESTAMP si la entrada es date + time.

TIMESTAMPTZ si la entrada es date + timetz.

### Ejemplos

#### Configuración de ejemplo

Para configurar las tablas TIME_TEST y TIMETZ_TEST utilizadas en los ejemplos, use el siguiente comando.

```
create table time_test(time_val time);

insert into time_test values
('20:00:00'),
('00:00:00.5550'),
('00:58:00');
```

```
create table timetz_test(timetz_val timetz);

insert into timetz_test values
('04:00:00+00'),
('00:00:00.5550+00'),
('05:58:00+00');
```

Ejemplos con una columna TIME

La siguiente tabla de ejemplo, TIME_TEST, tiene una columna TIME_VAL (tipo TIME) con tres valores insertados.

```
select time_val from time_test;

time_val
--------------------
20:00:00
00:00:00.5550
00:58:00
```

En el siguiente ejemplo, se concatena un literal de fecha y una columna TIME_VAL.

```
select date '2000-01-02' + time_val as ts from time_test;

ts
---------------------
2000-01-02 20:00:00
2000-01-02 00:00:00.5550
2000-01-02 00:58:00
```

En el siguiente ejemplo, se concatena un literal de fecha y un literal de hora.

```
select date '2000-01-01' + time '20:00:00' as ts;

         ts
---------------------
 2000-01-01 20:00:00
```

En el siguiente ejemplo se concatenan un literal de hora y un literal de fecha.

```
select time '20:00:00' + date '2000-01-01' as ts;
```

```
         ts
---------------------
 2000-01-01 20:00:00
```

Ejemplos con una columna TIMETZ

La siguiente tabla de ejemplo, TIMETZ_TEST, tiene una columna TIMETZ_VAL (tipo TIMETZ) con tres valores insertados.

```
select timetz_val from timetz_test;

timetz_val
------------------
04:00:00+00
00:00:00.5550+00
05:58:00+00
```

En el siguiente ejemplo, se concatena un literal de fecha y una columna TIMETZ_VAL.

```
select date '2000-01-01' + timetz_val as ts from timetz_test;
ts
---------------------
2000-01-01 04:00:00+00
2000-01-01 00:00:00.5550+00
2000-01-01 05:58:00+00
```

En el siguiente ejemplo se concatenan una columna TIMETZ_VAL y un literal de fecha.

```
select timetz_val + date '2000-01-01' as ts from timetz_test;
ts
---------------------
2000-01-01 04:00:00+00
2000-01-01 00:00:00.5550+00
2000-01-01 05:58:00+00
```

En el siguiente ejemplo, se concatena un literal de DATE y un literal de TIMETZ. El ejemplo devuelve un TIMESTAMPTZ que está en la zona horaria UTC de forma predeterminada. UTC está 8 horas por delante de PST, por lo que el resultado está 8 horas por delante de la hora de entrada.

```
select date '2000-01-01' + timetz '20:00:00 PST' as ts;
```

```
           ts
------------------------
 2000-01-02 04:00:00+00
```

## Función ADD_MONTHS

ADD_MONTHS agrega la cantidad de meses especificada a una expresión o un valor de fecha o marca temporal. La función DATEADD ofrece una funcionalidad similar.

### Sintaxis

```
ADD_MONTHS( {date | timestamp}, integer)
```

### Argumentos

date | timestamp

Una columna de tipo de datos `DATE` o `TIMESTAMP`, o una expresión que toma implícitamente el valor de un tipo `DATE` o `TIMESTAMP`. Si la fecha es el último día del mes, o si el mes resultante es más corto, la función devuelve el último día del mes en el resultado. Para otras fechas, el resultado tiene el mismo número de día que la expresión de fecha.

integer

Un valor de tipo de datos `INTEGER`. Use un número negativo para restar meses de las fechas.

### Tipo de retorno

TIMESTAMP

### Ejemplos

La siguiente consulta utiliza la función ADD_MONTHS dentro de una función TRUNC. La función TRUNC quita la hora del día del resultado de ADD_MONTHS. La función ADD_MONTHS agrega 12 meses a cada valor de la columna CALDATE. Los valores de la columna CALDATE son fechas.

```
select distinct trunc(add_months(caldate, 12)) as calplus12,
trunc(caldate) as cal
from date
order by 1 asc;

 calplus12  |    cal
```

```
------------+------------
 2009-01-01 | 2008-01-01
 2009-01-02 | 2008-01-02
 2009-01-03 | 2008-01-03
...
(365 rows)
```

En el siguiente ejemplo se utiliza la función ADD_MONTHS para sumar 1 mes a un valor timestamp.

```
select add_months('2008-01-01 05:07:30', 1);

add_months
---------------------
2008-02-01 05:07:30
```

En los ejemplos a continuación, se demuestra el comportamiento resultante cuando la función ADD_MONTHS opera sobre fechas con meses que tienen diferente cantidad de días. En este ejemplo se muestra cómo la función controla sumar 1 mes al 31 de marzo y sumar 1 mes al 30 de abril. Abril tiene 30 días, por lo que sumar 1 mes al 31 de marzo da como resultado el 30 de abril. Mayo tiene 31 días, por lo que sumar 1 mes al 30 de abril da como resultado el 31 de mayo.

```
select add_months('2008-03-31',1);

add_months
---------------------
2008-04-30 00:00:00

select add_months('2008-04-30',1);

add_months
---------------------
2008-05-31 00:00:00
```

## Función AT TIME ZONE

AT TIME ZONE especifica la zona horaria que debe usarse con una expresión TIMESTAMP o TIMESTAMPTZ.

### Sintaxis

```
AT TIME ZONE 'timezone'
```

Argumentos

timezone

La TIMEZONE para el valor devuelto. La zona horaria puede especificarse como un nombre (por ejemplo, **'Africa/Kampala'** o **'Singapore'**) o como una abreviatura (por ejemplo, **'UTC'** o **'PDT'**).

Para ver una lista de los nombres de zonas horarias compatibles, ejecute el siguiente comando.

```
select pg_timezone_names();
```

Para ver una lista de las abreviaturas de zonas horarias compatibles, ejecute el siguiente comando.

```
select pg_timezone_abbrevs();
```

Para obtener más información y ejemplos, consulta Notas de uso de la zona horaria.

Tipo de retorno

TIMESTAMPTZ, cuando se usa con una expresión TIMESTAMP. TIMESTAMP, cuando se usa con una expresión TIMESTAMPTZ.

Ejemplos

En el siguiente ejemplo, se convierte un valor de marca temporal sin zona horaria y se interpreta como hora MST (UTC+7 en POSIX). En el ejemplo se devuelve un valor del tipo de datos TIMESTAMPTZ para la zona horaria UTC. Si configura su zona horaria predeterminada en una zona horaria distinta de UTC, es posible que vea un resultado diferente.

```
SELECT TIMESTAMP '2001-02-16 20:38:40' AT TIME ZONE 'MST';

timezone
------------------------
2001-02-17 03:38:40+00
```

En el siguiente ejemplo, se toma una marca temporal de entrada con un valor de zona horaria cuya zona horaria especificada es EST (UTC+5 en POSIX) y se lo convierte a MST (UTC+7 en POSIX). En el ejemplo se devuelve un valor de tipo de datos TIMESTAMP.

```
SELECT TIMESTAMPTZ '2001-02-16 20:38:40-05' AT TIME ZONE 'MST';

timezone
------------------------
2001-02-16 18:38:40
```

## Función CONVERT_TIMEZONE

CONVERT_TIMEZONE convierte una marca temporal de una zona horaria a otra. La función se ajusta automáticamente al horario de verano.

### Sintaxis

```
CONVERT_TIMEZONE( ['source_timezone',] 'target_timezone', 'timestamp')
```

### Argumentos

source_timezone

(Opcional) La zona horaria de la marca temporal actual. El valor predeterminado es UTC. Para obtener más información, consulte Notas de uso de la zona horaria.

target_timezone

La zona horaria para la marca temporal nueva. Para obtener más información, consulte Notas de uso de la zona horaria.

timestamp

Una columna de marca temporal o una expresión que, implícitamente, se convierte en una marca temporal.

### Tipo de retorno

TIMESTAMP

### Notas de uso de la zona horaria

Tanto source_timezone como target_timezone pueden especificarse como un nombre de zona horaria (por ejemplo, “África/Kampala” o “Singapur”) o como una abreviatura de zona horaria (por ejemplo, “UTC” o “PDT”). No es necesario convertir los nombres de las zonas horarias en nombres

ni las abreviaturas en abreviaturas. Por ejemplo, puede elegir una marca de tiempo en el nombre de zona horaria de origen “Singapur” y convertirla en una marca de tiempo en la abreviatura de zona horaria “PDT”.

> **Note**
>
> Los resultados de usar un nombre de zona horaria o una abreviatura de zona horaria pueden ser diferentes debido a la hora estacional local (por ejemplo, el horario de verano).

Uso de un nombre de zona horaria

Para ver una lista actualizada y completa de los nombres de zonas horarias, ejecute el siguiente comando.

```
select pg_timezone_names();
```

Cada fila contiene una cadena separada por comas con el nombre de la zona horaria, la abreviatura, el desfase respecto de UTC y el indicador de si la zona horaria sigue el horario de verano (`t` o `f`). Por ejemplo, en el siguiente fragmento se muestran dos filas resultantes. La primera fila es la zona horaria `Antarctica/South Pole`, la abreviatura `NZDT`, con un desfase de `13:00:00` respecto de UTC y `f` para indicar que no sigue el horario de verano. La segunda fila es la zona horaria `Europe/Paris`, la abreviatura `CET`, con un desfase de `01:00:00` respecto de UTC y `f` para indicar que sigue el horario de verano.

```
pg_timezone_names
------------------
(Antarctica/South_Pole,NZDT,13:00:00,t)
(Europe/Paris,CET,01:00:00,f)
```

Ejecute la instrucción SQL para obtener la lista completa y buscar un nombre de zona horaria. Se devuelven aproximadamente 600 filas. Aunque algunos de los nombres de zonas horarias devueltos son siglas o acrónimos en mayúsculas (por ejemplo: GB, PRC, ROK), la función CONVERT_TIMEZONE los trata como nombres de zonas horarias, no abreviaturas de zonas horarias.

Si especifica una zona horaria mediante un nombre de zona horaria, CONVERT_TIMEZONE se ajusta automáticamente al horario de verano (DST) o a cualquier otro protocolo de temporada local, como el horario de verano, el horario estándar o el horario de invierno, que esté vigente en esa

zona horaria durante la fecha y hora especificadas por “timestamp”. Por ejemplo, “Europa/Londres” representa UTC en invierno y suma una hora en verano.

Uso de una abreviatura de zona horaria

Para ver una lista actualizada y completa de las abreviaturas de zona horaria, ejecute el siguiente comando.

```
select pg_timezone_abbrevs();
```

Los resultados contienen una cadena separada por comas con la abreviatura de la zona horaria, el desfase con respecto de UTC y un indicador de si la zona horaria sigue el horario de verano (`t` o `f`). Por ejemplo, en el siguiente fragmento se muestran dos filas resultantes. La primera fila contiene la abreviatura del horario de verano del Pacífico `PDT`, con un desfase de `-07:00:00` con respecto de UTC y `t` para indicar que sigue el horario de verano. La segunda fila contiene la abreviatura de hora estándar del Pacífico `PST`, con un desfase de `-08:00:00` con respecto de UTC y `f` para indicar que no sigue el horario de verano.

```
pg_timezone_abbrevs
--------------------
(PDT,-07:00:00,t)
(PST,-08:00:00,f)
```

Ejecute la instrucción SQL para obtener la lista completa y buscar una abreviatura basada en el indicador de desfase y horario de verano. Se devuelven aproximadamente 200 filas.

Las abreviaturas de zona horaria representan un desfase fijo a partir de UTC. Si especifica una zona horaria mediante una abreviatura de zona horaria, CONVERT_TIMEZONE usa el desfase fijo con respecto a UTC para cualquier protocolo de temporada local.

Uso del formato estilo POSIX

Una especificación de zona horaria de estilo POSIX toma la forma STDoffset o STDoffsetDST, donde STD es la abreviatura de la zona horaria, offset es el ajuste numérico en horas hacia el oeste desde UTC y DST es la abreviatura opcional del horario de verano de la zona. Se supone que el horario de verano adelanta una hora, en función del desfase dado.

Los formatos de zona horaria estilo POSIX usan desfases positivos al oeste de Greenwich, al contrario que la convención de la norma ISO 8601, que usa desfases positivos al este de Greenwich.

A continuación, se muestran algunos ejemplos de zonas horarias estilo POSIX:

- PST8
- PST8PDT
- EST5
- EST5EDT

> **Note**
>
> Amazon Redshift no valida las especificaciones de zona horaria con estilo POSIX, por lo que es posible establecer la zona horaria en un valor no válido. Por ejemplo, el siguiente comando no devuelve un mensaje de error, a pesar de que establece la zona horaria en un valor no válido.
>
> ```
> set timezone to ‘xxx36’;
> ```

Ejemplos

En muchos ejemplos se utiliza el conjunto de datos de muestra TICKIT. Para obtener más información, consulte Base de datos de muestra.

En el siguiente ejemplo, se convierte el valor de la marca temporal de la zona horaria UTC predeterminada a la zona horaria PST.

```
select convert_timezone('PST', '2008-08-21 07:23:54');

 convert_timezone
------------------------
2008-08-20 23:23:54
```

En el siguiente ejemplo, el valor de la marca temporal que aparece en la columna LISTTIME se convierte de la zona horaria UTC predeterminada a la zona horaria PST. Aunque la marca temporal se encuentra dentro del periodo de horario de verano, se convierte a horario estándar porque la zona horaria objetivo se especifica como una abreviatura (PST).

```
select listtime, convert_timezone('PST', listtime) from listing
where listid = 16;

     listtime       |   convert_timezone
```

```
--------------------+--------------------
2008-08-24 09:36:12     2008-08-24 01:36:12
```

En el siguiente ejemplo, la marca temporal que aparece en la columna LISTTIME se convierte de la zona horaria predeterminada UTC a la zona horaria US/Pacific. La zona horaria objetivo usa un nombre de zona horaria y la marca temporal se encuentra dentro del periodo de horario de verano, por lo que la función devuelve el horario de verano.

```
select listtime, convert_timezone('US/Pacific', listtime) from listing
where listid = 16;

     listtime       |   convert_timezone
--------------------+---------------------
2008-08-24 09:36:12 | 2008-08-24 02:36:12
```

En el siguiente ejemplo, se convierte una cadena de marca temporal de EST a PST:

```
select convert_timezone('EST', 'PST', '20080305 12:25:29');

 convert_timezone
-------------------
2008-03-05 09:25:29
```

En el siguiente ejemplo, se convierte una marca temporal al horario del este de Estados Unidos estándar porque la zona horaria objetivo usa un nombre de zona horaria (America/New_York) y la marca temporal se encuentra dentro del periodo estándar.

```
select convert_timezone('America/New_York', '2013-02-01 08:00:00');

 convert_timezone
---------------------
2013-02-01 03:00:00
(1 row)
```

En el siguiente ejemplo, se convierte la marca temporal al horario de verano del este de Estados Unidos porque la zona horaria objetivo usa un nombre de zona horaria (America/New_York) y la marca temporal se encuentra dentro del periodo de horario de verano.

```
select convert_timezone('America/New_York', '2013-06-01 08:00:00');

 convert_timezone
```

```
---------------------
2013-06-01 04:00:00
(1 row)
```

En el siguiente ejemplo, se demuestra el uso de desplazamientos.

```
SELECT CONVERT_TIMEZONE('GMT','NEWZONE +2','2014-05-17 12:00:00') as newzone_plus_2,
CONVERT_TIMEZONE('GMT','NEWZONE-2:15','2014-05-17 12:00:00') as newzone_minus_2_15,
CONVERT_TIMEZONE('GMT','America/Los_Angeles+2','2014-05-17 12:00:00') as la_plus_2,
CONVERT_TIMEZONE('GMT','GMT+2','2014-05-17 12:00:00') as gmt_plus_2;

   newzone_plus_2    | newzone_minus_2_15  |      la_plus_2      |     gmt_plus_2
---------------------+---------------------+---------------------+---------------------
2014-05-17 10:00:00 | 2014-05-17 14:15:00 | 2014-05-17 10:00:00 | 2014-05-17 10:00:00
(1 row)
```

## Función CURRENT_DATE

CURRENT_DATE devuelve una fecha en la zona horaria de la sesión actual (que es UTC de manera predeterminada) en el formato predeterminado: AAAA-MM-DD.

> **Note**
>
> CURRENT_DATE devuelve la fecha de comienzo de la transacción actual, no de la instrucción actual. Pensemos en el escenario en el que se inicia una transacción con varias instrucciones el 10/01/08 a las 23:59 y la instrucción que contiene CURRENT_DATE se ejecuta el 10/02/08 a las 00:00. CURRENT_DATE devuelve `10/01/08`, no `10/02/08`.

### Sintaxis

```
CURRENT_DATE
```

### Tipo de retorno

DATE

### Ejemplos

En el siguiente ejemplo, se devuelve la fecha actual (en las Región de AWS donde se ejecuta la función).

```
select current_date;

   date
------------
2008-10-01
```

En el ejemplo siguiente se crea una tabla, se inserta una fila en la que el valor predeterminado de la columna `todays_date` es CURRENT_DATE y, a continuación, se seleccionan todas las filas de la tabla.

```
CREATE TABLE insert_dates(
    label varchar(128) NOT NULL,
    todays_date DATE DEFAULT CURRENT_DATE);

INSERT INTO insert_dates(label)
VALUES('Date row inserted');

SELECT * FROM insert_dates;


 label            | todays_date
------------------+-------------
Date row inserted | 2023-05-10
```

## Función DATE_CMP

DATE_CMP compara dos fechas. La función devuelve `0` si las fechas son idénticas, `1` si date1 es posterior y `-1` si date2 es posterior.

### Sintaxis

```
DATE_CMP(date1, date2)
```

### Argumentos

#### date1

Una columna de tipo de datos `DATE` o una expresión que toma el valor de un tipo `DATE`.

date2

Una columna de tipo de datos DATE o una expresión que toma el valor de un tipo DATE.

Tipo de retorno

INTEGER

Ejemplos

La siguiente consulta compara los valores DATE de la columna CALDATE con la fecha 4 de enero del 2008 y devuelve si el valor en CALDATE es anterior (-1), igual (0) o posterior (1) al 4 de enero del 2008:

```
select caldate, '2008-01-04',
date_cmp(caldate,'2008-01-04')
from date
order by dateid
limit 10;

 caldate   |  ?column?  | date_cmp
-----------+------------+----------
2008-01-01 | 2008-01-04 |       -1
2008-01-02 | 2008-01-04 |       -1
2008-01-03 | 2008-01-04 |       -1
2008-01-04 | 2008-01-04 |        0
2008-01-05 | 2008-01-04 |        1
2008-01-06 | 2008-01-04 |        1
2008-01-07 | 2008-01-04 |        1
2008-01-08 | 2008-01-04 |        1
2008-01-09 | 2008-01-04 |        1
2008-01-10 | 2008-01-04 |        1
(10 rows)
```

## Función DATE_CMP_TIMESTAMP

DATE_CMP_TIMESTAMP compara una fecha con una marca temporal y devuelve 0 si los valores son idénticos, 1 si date es posterior cronológicamente y -1 si timestamp es posterior.

Sintaxis

```
DATE_CMP_TIMESTAMP(date, timestamp)
```

### Argumentos

date

Una columna de tipo de datos DATE o una expresión que toma el valor de un tipo DATE.

timestamp

Una columna de tipo de datos TIMESTAMP o una expresión que toma el valor de un tipo TIMESTAMP.

### Tipo de retorno

INTEGER

### Ejemplos

En el siguiente ejemplo, se compara la fecha 2008-06-18 con LISTTIME. Los valores de la columna LISTTIME son marcas temporales. Las listas hechas antes de esa fecha devuelven 1; las hechas después de ella, -1.

```
select listid, '2008-06-18', listtime,
date_cmp_timestamp('2008-06-18', listtime)
from listing
order by 1, 2, 3, 4
limit 10;

 listid |  ?column?  |      listtime       | date_cmp_timestamp
--------+------------+---------------------+--------------------
      1 | 2008-06-18 | 2008-01-24 06:43:29 |                  1
      2 | 2008-06-18 | 2008-03-05 12:25:29 |                  1
      3 | 2008-06-18 | 2008-11-01 07:35:33 |                 -1
      4 | 2008-06-18 | 2008-05-24 01:18:37 |                  1
      5 | 2008-06-18 | 2008-05-17 02:29:11 |                  1
      6 | 2008-06-18 | 2008-08-15 02:08:13 |                 -1
      7 | 2008-06-18 | 2008-11-15 09:38:15 |                 -1
      8 | 2008-06-18 | 2008-11-09 05:07:30 |                 -1
      9 | 2008-06-18 | 2008-09-09 08:03:36 |                 -1
     10 | 2008-06-18 | 2008-06-17 09:44:54 |                  1
(10 rows)
```

## Función DATE_CMP_TIMESTAMPTZ

DATE_CMP_TIMESTAMPTZ compara una fecha con una marca temporal con zona horaria y devuelve `0` si los valores son idénticos, `1` si date es posterior cronológicamente y `-1` si timestamptz es posterior.

### Sintaxis

```
DATE_CMP_TIMESTAMPTZ(date, timestamptz)
```

### Argumentos

date

Una columna de tipo de datos `DATE` o una expresión que toma implícitamente el valor de un tipo `DATE`.

timestamptz

Una columna de tipo de datos `TIMESTAMPTZ` o una expresión que toma implícitamente el valor de un tipo `TIMESTAMPTZ`.

### Tipo de retorno

INTEGER

### Ejemplos

En el siguiente ejemplo, se compara la fecha `2008-06-18` con LISTTIME. Las listas hechas antes de esa fecha devuelven 1; las hechas después de ella, `-1`.

```
select listid, '2008-06-18', CAST(listtime AS timestamptz),
date_cmp_timestamptz('2008-06-18', CAST(listtime AS timestamptz))
from listing
order by 1, 2, 3, 4
limit 10;

 listid |  ?column?  |      timestamptz       | date_cmp_timestamptz
--------+------------+------------------------+----------------------
      1 | 2008-06-18 | 2008-01-24 06:43:29+00 |                    1
      2 | 2008-06-18 | 2008-03-05 12:25:29+00 |                    1
      3 | 2008-06-18 | 2008-11-01 07:35:33+00 |                   -1
```

```
      4 | 2008-06-18 | 2008-05-24 01:18:37+00 |                  1
      5 | 2008-06-18 | 2008-05-17 02:29:11+00 |                  1
      6 | 2008-06-18 | 2008-08-15 02:08:13+00 |                 -1
      7 | 2008-06-18 | 2008-11-15 09:38:15+00 |                 -1
      8 | 2008-06-18 | 2008-11-09 05:07:30+00 |                 -1
      9 | 2008-06-18 | 2008-09-09 08:03:36+00 |                 -1
     10 | 2008-06-18 | 2008-06-17 09:44:54+00 |                  1
(10 rows)
```

## Función DATEADD

Aumenta un valor de DATE, TIME, TIMETZ o TIMESTAMP según un intervalo especificado.

### Sintaxis

```
DATEADD( datepart, interval, {date|time|timetz|timestamp} )
```

### Argumentos

datepart

La parte de la fecha (año, mes, día u hora, por ejemplo) sobre la que opera la función. Para obtener más información, consulte Partes de fecha para funciones de fecha o marca temporal.

interval

Un número entero que especifica el intervalo (cantidad de días, por ejemplo) por agregar a la expresión objetivo. Un número entero negativo resta al intervalo.

date|time|timetz|timestamp

Una columna DATE, TIME, TIMETZ o TIMESTAMP o una expresión que, de forma implícita, se convierte a una DATE, TIME, TIMETZ o TIMESTAMP. La expresión de DATE, TIME, TIMETZ o TIMESTAMP debe contener la parte de fecha especificada.

### Tipo de retorno

TIMESTAMP, TIME o TIMETZ según el tipo de datos de entrada.

### Ejemplos con una columna DATE

En el siguiente ejemplo, se agregan 30 días a cada fecha en noviembre que se encuentra en la tabla de DATE.

```
select dateadd(day,30,caldate) as novplus30
from date
where month='NOV'
order by dateid;

novplus30
---------------------
2008-12-01 00:00:00
2008-12-02 00:00:00
2008-12-03 00:00:00
...
(30 rows)
```

En el siguiente ejemplo, se agregan 18 meses a un valor de fecha literal.

```
select dateadd(month,18,'2008-02-28');

date_add
---------------------
2009-08-28 00:00:00
(1 row)
```

El nombre predeterminado de la columna para una función DATEADD es DATE_ADD. La marca temporal predeterminada para un valor de fecha es `00:00:00`.

En el siguiente ejemplo, se agregan 30 minutos a un valor de fecha que no especifica una marca temporal.

```
select dateadd(m,30,'2008-02-28');

date_add
---------------------
2008-02-28 00:30:00
(1 row)
```

Puede nombrar las partes de la fecha de manera completa o abreviada. En este caso, m representa a los minutos y no a lo meses.

Ejemplos con una columna TIME

La siguiente tabla de ejemplo, TIME_TEST, tiene una columna TIME_VAL (tipo TIME) con tres valores insertados.

```
select time_val from time_test;

time_val
---------------------
20:00:00
00:00:00.5550
00:58:00
```

En el siguiente ejemplo, se agregan 5 minutos a cada TIME_VAL de la tabla TIME_TEST.

```
select dateadd(minute,5,time_val) as minplus5 from time_test;

minplus5
---------------
20:05:00
00:05:00.5550
01:03:00
```

En el siguiente ejemplo, se agregan 8 horas a un valor de fecha literal.

```
select dateadd(hour, 8, time '13:24:55');

date_add
---------------
21:24:55
```

El siguiente ejemplo muestra cuando una hora supera las 24:00:00 o cuando es antes de las 00:00:00.

```
select dateadd(hour, 12, time '13:24:55');

date_add
---------------
01:24:55
```

Ejemplos con una columna TIMETZ

Los valores de salida en estos ejemplos están en formato UTC, que es la zona horaria predeterminada.

La siguiente tabla de ejemplo, TIMETZ_TEST, tiene una columna TIMETZ_VAL (tipo TIMETZ) con tres valores insertados.

```
select timetz_val from timetz_test;

timetz_val
------------------
04:00:00+00
00:00:00.5550+00
05:58:00+00
```

En el siguiente ejemplo, se agregan 5 minutos a cada TIMETZ_VAL de la tabla TIMETZ_TEST.

```
select dateadd(minute,5,timetz_val) as minplus5_tz from timetz_test;

minplus5_tz
---------------
04:05:00+00
00:05:00.5550+00
06:03:00+00
```

En el siguiente ejemplo, se agregan 2 horas a un valor de timetz literal.

```
select dateadd(hour, 2, timetz '13:24:55 PST');

date_add
---------------
23:24:55+00
```

Ejemplos con una columna TIMESTAMP

Los valores de salida en estos ejemplos están en formato UTC, que es la zona horaria predeterminada.

La siguiente tabla de ejemplo, TIMESTAMP_TEST tiene una columna TIMESTAMP_VAL (tipo TIMESTAMP) con tres valores insertados.

```
SELECT timestamp_val FROM timestamp_test;

timestamp_val
```

```
------------------
1988-05-15 10:23:31
2021-03-18 17:20:41
2023-06-02 18:11:12
```

El siguiente ejemplo agrega 20 años solo a los valores TIMESTAMP_VAL en TIMESTAMP_TEST anteriores al año 2000.

```
SELECT dateadd(year,20,timestamp_val)
FROM timestamp_test
WHERE timestamp_val < to_timestamp('2000-01-01 00:00:00', 'YYYY-MM-DD HH:MI:SS');

date_add
---------------
2008-05-15 10:23:31
```

En el siguiente ejemplo se agregan 5 segundos a un valor de marca temporal literal escrito sin un indicador de segundos.

```
SELECT dateadd(second, 5, timestamp '2001-06-06');

date_add
---------------
2001-06-06 00:00:05
```

Notas de uso

Las funciones DATEADD(month, ...) y ADD_MONTHS administran las fechas que caen en fin de mes de manera diferente:

- ADD_MONTHS: si la fecha que se agrega es el último día del mes, el resultado es siempre el último día del mes en el resultado, sin importar su longitud. Por ejemplo, 30 de abril + 1 mes es 31 de mayo.

  ```
  select add_months('2008-04-30',1);

  add_months
  ---------------------
  2008-05-31 00:00:00
  (1 row)
  ```

- DATEADD: si en la fecha que se agrega hay menos días que en el mes del resultado, el resultado será el día correspondiente del mes del resultado, no el último día de ese mes. Por ejemplo, 30 de abril + 1 mes es 30 de mayo.

```
select dateadd(month,1,'2008-04-30');

date_add
---------------------
2008-05-30 00:00:00
(1 row)
```

La función DATEADD administra la fecha 02-29 de año bisiesto de manera diferente con dateadd (month, 12,…) o dateadd (year, 1, …).

```
select dateadd(month,12,'2016-02-29');

date_add
---------------------
2017-02-28 00:00:00

select dateadd(year, 1, '2016-02-29');

date_add
---------------------
2017-03-01 00:00:00
```

## Función DATEDIFF

DATEDIFF devuelve la diferencia entre las partes de fecha de dos expresiones de fecha u hora.

Sintaxis

```
DATEDIFF( datepart, {date|time|timetz|timestamp}, {date|time|timetz|timestamp} )
```

Argumentos

datepart

La parte específica del valor de fecha u hora (año, mes o día; hora, minuto, segundo, milisegundo o microsegundo) sobre la que opera la función. Para obtener más información, consulte Partes de fecha para funciones de fecha o marca temporal.

En concreto, DATEDIFF determina la cantidad de límites de la parte de fecha que se cruzan entre las dos expresiones. Suponga, por ejemplo, que calcula la diferencia en años entre dos fechas, `12-31-2008` y `01-01-2009`. En este caso, la función devuelve 1 año a pesar de que estas fechas solo tienen un día de diferencia. Si busca la diferencia en horas entre dos marcas temporales, `01-01-2009 8:30:00` y `01-01-2009 10:00:00`, el resultado es 2 horas. Si busca la diferencia en horas entre dos marcas temporales, `8:30:00` y `10:00:00`, el resultado es 2 horas.

date|time|timetz|timestamp

Una columna o expresiones de DATE, TIME, TIMETZ o TIMESTAMP que se convierten de forma implícita en DATE, TIME, TIMETZ o TIMESTAMP. Ambas expresiones deben contener la parte de fecha u hora especificada. Si la segunda fecha u hora es posterior a la primera, el resultado es positivo. Si la segunda fecha u hora es anterior a la primera, el resultado es negativo.

Tipo de retorno

BIGINT

Ejemplos con una columna DATE

En el siguiente ejemplo, se encuentra la diferencia, en cantidad de semanas, entre dos valores de fecha literales.

```
select datediff(week,'2009-01-01','2009-12-31') as numweeks;

numweeks
----------
52
(1 row)
```

En el siguiente ejemplo, se encuentra la diferencia, en horas, entre dos valores de fecha literales. Cuando no se proporciona el valor de la hora para una fecha, de forma predeterminada es 00:00:00.

```
select datediff(hour, '2023-01-01', '2023-01-03 05:04:03');

date_diff
----------
53
(1 row)
```

En el siguiente ejemplo se encuentra la diferencia, en días, entre dos valores TIMESTAMETZ literales.

```
Select datediff(days, 'Jun 1,2008  09:59:59 EST', 'Jul 4,2008  09:59:59 EST')

date_diff
----------
33
```

En el siguiente ejemplo, se encuentra la diferencia, en días, entre dos fechas de la misma fila de una tabla.

```
select * from date_table;

start_date |   end_date
-----------+-----------
2009-01-01 | 2009-03-23
2023-01-04 | 2024-05-04
(2 rows)

select datediff(day, start_date, end_date) as duration from date_table;

duration
---------
      81
     486
(2 rows)
```

En el siguiente ejemplo, se encuentra la diferencia, en cantidad de trimestres, entre un valor literal del pasado y la fecha de hoy. En este ejemplo, se asume que la fecha actual es 5 de junio del 2008. Puede nombrar las partes de la fecha de manera completa o abreviada. El nombre predeterminado de la columna para la función DATEDIFF es DATE_DIFF.

```
select datediff(qtr, '1998-07-01', current_date);
```

```
date_diff
-----------
40
(1 row)
```

En este ejemplo, se unen las tablas SALES y LISTING para calcular cuántos días después de indicarse se vendieron los tickets de los listados 1000 a 1005. La espera más prolongada para la venta de estos listados fue de 15 días, y la más corta, de menos de 1 día (0 días).

```
select priceperticket,
datediff(day, listtime, saletime) as wait
from sales, listing where sales.listid = listing.listid
and sales.listid between 1000 and 1005
order by wait desc, priceperticket desc;

priceperticket | wait
---------------+------
 96.00         |   15
 123.00        |   11
 131.00        |    9
 123.00        |    6
 129.00        |    4
 96.00         |    4
 96.00         |    0
(7 rows)
```

En este ejemplo, se calculan las horas promedio esperadas por los vendedores para todas las ventas de tickets.

```
select avg(datediff(hours, listtime, saletime)) as avgwait
from sales, listing
where sales.listid = listing.listid;

avgwait
---------
465
(1 row)
```

### Ejemplos con una columna TIME

La siguiente tabla de ejemplo, TIME_TEST, tiene una columna TIME_VAL (tipo TIME) con tres valores insertados.

```
select time_val from time_test;

time_val
---------------------
20:00:00
00:00:00.5550
00:58:00
```

En el siguiente ejemplo, se encuentra la diferencia en cantidad de horas entre la columna TIME_VAL y un literal de tiempo.

```
select datediff(hour, time_val, time '15:24:45') from time_test;

 date_diff
-----------
        -5
        15
        15
```

En el siguiente ejemplo, se encuentra la diferencia en cantidad de minutos entre dos valores de tiempo literales.

```
select datediff(minute, time '20:00:00', time '21:00:00') as nummins;

nummins
----------
60
```

### Ejemplos con una columna TIMETZ

La siguiente tabla de ejemplo, TIMETZ_TEST, tiene una columna TIMETZ_VAL (tipo TIMETZ) con tres valores insertados.

```
select timetz_val from timetz_test;

timetz_val
```

```
------------------
04:00:00+00
00:00:00.5550+00
05:58:00+00
```

En el siguiente ejemplo, se encuentran las diferencias en la cantidad de horas, entre un literal TIMETZ y timetz_val.

```
select datediff(hours, timetz '20:00:00 PST', timetz_val) as numhours from timetz_test;

numhours
----------
0
-4
1
```

En el siguiente ejemplo, se encuentra la diferencia en cantidad de horas entre dos valores TIMETZ literales.

```
select datediff(hours, timetz '20:00:00 PST', timetz '00:58:00 EST') as numhours;

numhours
----------
1
```

## Función DATE_PART

DATE_PART extrae los valores de parte de fecha a partir de una expresión. DATE_PART es sinónimo de la función PGDATE_PART.

### Sintaxis

```
DATE_PART(datepart, {date|timestamp})
```

### Argumentos

#### datepart

Un literal o cadena de identificación de la parte específica del valor de la fecha (por ejemplo, año, mes o día) en la que actúa la función. Para obtener más información, consulte Partes de fecha para funciones de fecha o marca temporal.

{date|timestamp}

Una columna de fecha, una columna de marca de tiempo o una expresión que se convierte implícitamente en una fecha o una marca de tiempo. La columna o expresión en fecha o marca de tiempo debe contener la parte de fecha especificada en parte de fecha.

Tipo de retorno

DOBLE

Ejemplos

El nombre predeterminado de la columna para la función DATE_PART es `pgdate_part`.

Para obtener más información acerca de los datos utilizados en algunos de estos ejemplos, consulte Base de datos de muestra.

En el ejemplo siguiente se encuentra el minuto de un literal de marca temporal.

```
SELECT DATE_PART(minute, timestamp '20230104 04:05:06.789');

pgdate_part
-----------
          5
```

En el ejemplo siguiente, se encuentra el número de semana de un literal de marca temporal. El cálculo del número de semana sigue el estándar ISO 8601. Para obtener más información, consulte ISO 8601 en Wikipedia.

```
SELECT DATE_PART(week, timestamp '20220502 04:05:06.789');

pgdate_part
-----------
         18
```

En el ejemplo siguiente se encuentra el día del mes de un literal de marca temporal.

```
SELECT DATE_PART(day, timestamp '20220502 04:05:06.789');

pgdate_part
```

```
-----------
          2
```

En el ejemplo siguiente se encuentra el día de la semana de un literal de marca temporal. El cálculo del número del día de la semana es un número entero de 0 a 6, que empieza el domingo.

```
SELECT DATE_PART(dayofweek, timestamp '20220502 04:05:06.789');

pgdate_part
-----------
          1
```

En el ejemplo siguiente se encuentra el siglo de un literal de marca temporal. El cálculo del siglo sigue el estándar ISO 8601. Para obtener más información, consulte ISO 8601 en Wikipedia.

```
SELECT DATE_PART(century, timestamp '20220502 04:05:06.789');

pgdate_part
-----------
         21
```

En el ejemplo siguiente, se encuentra el milenio de un literal de marca de tiempo. El cálculo del milenio sigue el estándar ISO 8601. Para obtener más información, consulte ISO 8601 en Wikipedia.

```
SELECT DATE_PART(millennium, timestamp '20220502 04:05:06.789');

pgdate_part
-----------
          3
```

En el ejemplo siguiente, se encuentran los microsegundos de un literal de marca de tiempo. El cálculo de los microsegundos sigue el estándar ISO 8601. Para obtener más información, consulte ISO 8601 en Wikipedia.

```
SELECT DATE_PART(microsecond, timestamp '20220502 04:05:06.789');

pgdate_part
-----------
    789000
```

En el ejemplo siguiente se encuentra el mes de un literal de fecha.

```
SELECT DATE_PART(month, date '20220502');

pgdate_part
-----------
          5
```

En el siguiente ejemplo, se aplica la función DATE_PART a una columna de una tabla.

```
SELECT date_part(w, listtime) AS weeks, listtime
FROM listing
WHERE listid=10


weeks |      listtime
------+---------------------
 25   | 2008-06-17 09:44:54
(1 row)
```

Puede especificar las partes de fecha utilizando su nombre completo o una abreviatura; en este caso, w quiere decir semanas.

La parte de fecha de día de la semana devuelve un número entero de 0 a 6, comenzando por domingo. Use DATE_PART con dow (DAYOFWEEK) para visualizar los eventos que se llevan a cabo un sábado.

```
SELECT date_part(dow, starttime) AS dow, starttime
FROM event
WHERE date_part(dow, starttime)=6
ORDER BY 2,1;

 dow |      starttime
-----+---------------------
   6 | 2008-01-05 14:00:00
   6 | 2008-01-05 14:00:00
   6 | 2008-01-05 14:00:00
   6 | 2008-01-05 14:00:00
...
(1147 rows)
```

## Función DATE_PART_YEAR

La función DATE_PART_YEAR extrae el año de una fecha.

### Sintaxis

```
DATE_PART_YEAR(date)
```

### Argumento

date

Una columna de tipo de datos DATE o una expresión que toma implícitamente el valor de un tipo DATE.

### Tipo de retorno

INTEGER

### Ejemplos

En el ejemplo siguiente se encuentra el año de un literal de fecha.

```
SELECT DATE_PART_YEAR(date '20220502 04:05:06.789');

date_part_year
---------------
2022
```

En el siguiente ejemplo, se extrae el año de la columna CALDATE. Los valores de la columna CALDATE son fechas. Para obtener más información sobre los datos que se utilizan en este ejemplo, consulte Base de datos de muestra.

```
select caldate, date_part_year(caldate)
from date
order by
dateid limit 10;

 caldate   | date_part_year
-----------+----------------
2008-01-01 |           2008
```

```
2008-01-02 |           2008
2008-01-03 |           2008
2008-01-04 |           2008
2008-01-05 |           2008
2008-01-06 |           2008
2008-01-07 |           2008
2008-01-08 |           2008
2008-01-09 |           2008
2008-01-10 |           2008
(10 rows)
```

## Función DATE_TRUNC

La función DATE_TRUNC trunca todo literal o expresión de marca temporal basado en la parte de fecha especificada, como la hora, la semana o el mes.

### Sintaxis

```
DATE_TRUNC('datepart', timestamp)
```

### Argumentos

datepart

La parte de fecha que trunca el valor de una marca temporal. La timestamp de entrada se trunca en la precisión de la entrada datepart. Por ejemplo, `month` se trunca en el primer día del mes. Los formatos válidos son los siguientes:

- microsegundo, microsegundos
- milisegundo, milisegundos
- segundo, segundos
- minuto, minutos
- hora, horas
- día, días
- semana, semanas
- mes, meses
- trimestre, trimestres
- año, años

- década, décadas
- siglo, siglos
- milenio, milenios

Para obtener más información sobre las abreviaturas de algunos formatos, consulte Partes de fecha para funciones de fecha o marca temporal.

timestamp

Una columna de marca temporal o una expresión que, implícitamente, se convierte en una marca temporal.

Tipo de retorno

TIMESTAMP

Ejemplos

Trunque la marca temporal de entrada en el segundo.

```
SELECT DATE_TRUNC('second', TIMESTAMP '20200430 04:05:06.789');
date_trunc
2020-04-30 04:05:06
```

Trunque la marca temporal de entrada al minuto.

```
SELECT DATE_TRUNC('minute', TIMESTAMP '20200430 04:05:06.789');
date_trunc
2020-04-30 04:05:00
```

Trunque la marca temporal de entrada a la hora.

```
SELECT DATE_TRUNC('hour', TIMESTAMP '20200430 04:05:06.789');
date_trunc
2020-04-30 04:00:00
```

Trunque la marca temporal de entrada al día.

```
SELECT DATE_TRUNC('day', TIMESTAMP '20200430 04:05:06.789');
date_trunc
```

```
2020-04-30 00:00:00
```

Trunque la marca temporal de entrada al primer día de un mes.

```
SELECT DATE_TRUNC('month', TIMESTAMP '20200430 04:05:06.789');
date_trunc
2020-04-01 00:00:00
```

Trunque la marca temporal de entrada al primer día de un trimestre.

```
SELECT DATE_TRUNC('quarter', TIMESTAMP '20200430 04:05:06.789');
date_trunc
2020-04-01 00:00:00
```

Trunque la marca temporal de entrada al primer día de un año.

```
SELECT DATE_TRUNC('year', TIMESTAMP '20200430 04:05:06.789');
date_trunc
2020-01-01 00:00:00
```

Trunque la marca temporal de entrada al primer día de un siglo.

```
SELECT DATE_TRUNC('millennium', TIMESTAMP '20200430 04:05:06.789');
date_trunc
2001-01-01 00:00:00
```

Trunque la marca temporal de entrada al lunes de un mes.

```
select date_trunc('week', TIMESTAMP '20220430 04:05:06.789');
date_trunc
2022-04-25 00:00:00
```

En el siguiente ejemplo, la función DATE_TRUNC usa la parte de fecha “week” (semana) para devolver la fecha del lunes de cada semana.

```
select date_trunc('week', saletime), sum(pricepaid) from sales where
saletime like '2008-09%' group by date_trunc('week', saletime) order by 1;

date_trunc  |    sum
------------+-------------
```

```
2008-09-01  | 2474899
2008-09-08  | 2412354
2008-09-15  | 2364707
2008-09-22  | 2359351
2008-09-29  |  705249
```

## Función EXTRACT

La función EXTRACT devuelve una parte de fecha u hora de un valor TIMESTAMP, TIMESTAMPTZ, TIME, TIMETZ, INTERVAL YEAR TO MONTH o INTERVAL DAY TO SECOND. Algunos ejemplos son día, mes, año, hora, minuto, segundo, milisegundo o microsegundo de una marca de tiempo.

### Sintaxis

```
EXTRACT(datepart FROM source)
```

### Argumentos

datepart

El subcampo de una fecha u hora que se va a extraer, como día, mes, año, hora, minuto, segundo, milisegundo o microsegundo. Para obtener los valores posibles, consulte Partes de fecha para funciones de fecha o marca temporal.

origen

Una columna o una expresión que se evalúa como un tipo de datos de TIMESTAMP, TIMESTAMPTZ, TIME, TIMETZ, INTERVAL YEAR TO MONTH o INTERVAL DAY TO SECOND.

### Tipo de retorno

INTEGER si el valor de origen se evalúa como tipo de datos TIMESTAMP, TIME, TIMETZ, INTERVAL YEAR TO MONTH o INTERVAL DAY TO SECOND.

DOUBLE PRECISION si el valor de origen se evalúa como el tipo de datos TIMESTAMPTZ.

### Ejemplos con TIMESTAMP

En el siguiente ejemplo, se determinan los números de las semanas para las ventas en las que el precio pagado fue 10 000 USD o más. En este ejemplo, se utilizan los datos de TICKIT. Para obtener más información, consulte Base de datos de muestra.

```
select salesid, extract(week from saletime) as weeknum
from sales
where pricepaid > 9999
order by 2;

salesid | weeknum
--------+---------
 159073 |       6
 160318 |       8
 161723 |      26
```

En el siguiente ejemplo, se devuelve el valor de minutos de un valor de marca temporal literal.

```
select extract(minute from timestamp '2009-09-09 12:08:43');

date_part
-----------
8
```

En el siguiente ejemplo, se devuelve el valor de milisegundos de un valor timestamp literal.

```
select extract(ms from timestamp '2009-09-09 12:08:43.101');

date_part
-----------
101
```

### Ejemplos con TIMESTAMPTZ

En el siguiente ejemplo, se devuelve el valor de año de un valor timestamptz literal.

```
select extract(year from timestamptz '1.12.1997 07:37:16.00 PST');

date_part
-----------
1997
```

### Ejemplos con TIME

La siguiente tabla de ejemplo, TIME_TEST, tiene una columna TIME_VAL (tipo TIME) con tres valores insertados.

```
select time_val from time_test;

time_val
---------------------
20:00:00
00:00:00.5550
00:58:00
```

En el siguiente ejemplo, se extraen los minutos de cada time_val.

```
select extract(minute from time_val) as minutes from time_test;

minutes
-----------
         0
         0
         58
```

En el siguiente ejemplo, se extraen las horas de cada time_val.

```
select extract(hour from time_val) as hours from time_test;

hours
-----------
         20
         0
         0
```

En el siguiente ejemplo, se extraen los milisegundos de un valor literal.

```
select extract(ms from time '18:25:33.123456');

 date_part
-----------
     123
```

Ejemplos con TIMETZ

La siguiente tabla de ejemplo, TIMETZ_TEST, tiene una columna TIMETZ_VAL (tipo TIMETZ) con tres valores insertados.

```
select timetz_val from timetz_test;

timetz_val
------------------
04:00:00+00
00:00:00.5550+00
05:58:00+00
```

En el siguiente ejemplo, se extraen las horas de cada timetz_val.

```
select extract(hour from timetz_val) as hours from time_test;

hours
-----------
         4
         0
         5
```

En el siguiente ejemplo, se extraen los milisegundos de un valor literal. Los literales no se convierten a UTC antes de que se procese la extracción.

```
select extract(ms from timetz '18:25:33.123456 EST');

 date_part
-----------
     123
```

En el siguiente ejemplo, se devuelve la hora de desfase horario respecto a UTC a partir de un valor timetz literal.

```
select extract(timezone_hour from timetz '1.12.1997 07:37:16.00 PDT');

date_part
-----------
-7
```

Ejemplos con INTERVAL YEAR TO MONTH e INTERVAL DAY TO SECOND

En el siguiente ejemplo, se extrae la parte del día de 1 del INTERVAL DAY TO SECOND que define 36 horas, es decir, 1 día y 12 horas.

```
select EXTRACT('days' from INTERVAL '36 hours' DAY TO SECOND)

 date_part
------------------
 1
```

En el siguiente ejemplo, se extrae la parte del mes de 3 del YEAR TO MONTH que define 15 meses, es decir, 1 año y 3 meses.

```
select EXTRACT('month' from INTERVAL '15 months' YEAR TO MONTH)

 date_part
------------------
 3
```

En el siguiente ejemplo, se extrae la parte del mes de 6 de 30 meses que es 2 años y 6 meses.

```
select EXTRACT('month' from INTERVAL '30' MONTH)

 date_part
------------------
 6
```

En el siguiente ejemplo, se extrae la parte de la hora de 2 de 50 horas que es 2 días y 2 horas.

```
select EXTRACT('hours' from INTERVAL '50' HOUR)

 date_part
------------------
 2
```

En el siguiente ejemplo, se extrae la parte de minutos de 11 de 1 hora, 11 minutos y 11 123 segundos.

```
select EXTRACT('minute' from INTERVAL '70 minutes 70.123 seconds' MINUTE TO SECOND)

 date_part
------------------
 11
```

En el siguiente ejemplo, se extrae la parte de segundos de 1.11 de 1 día, 1 hora, 1 minuto y 1,11 segundos.

```
select EXTRACT('seconds' from INTERVAL '1 day 1:1:1.11' DAY TO SECOND)

 date_part
------------------
 1.11
```

En el siguiente ejemplo, se extrae el número total de horas de un INTERVAL. Cada parte se extrae y se agrega a un total.

```
select EXTRACT('days' from INTERVAL '50' HOUR) * 24 + EXTRACT('hours' from INTERVAL
 '50' HOUR)

 ?column?
------------------
 50
```

En el siguiente ejemplo, se extrae el número total de segundos de un INTERVALO. Cada parte se extrae y se agrega a un total.

```
select EXTRACT('days' from INTERVAL '1 day 1:1:1.11' DAY TO SECOND) * 86400 +
       EXTRACT('hours' from INTERVAL '1 day 1:1:1.11' DAY TO SECOND) * 3600 +
       EXTRACT('minutes' from INTERVAL '1 day 1:1:1.11' DAY TO SECOND) * 60 +
       EXTRACT('seconds' from INTERVAL '1 day 1:1:1.11' DAY TO SECOND)

 ?column?
------------------
 90061.11
```

## Función GETDATE

GETDATE devuelve la fecha y hora actual en la zona horaria de la sesión actual (que es UTC de manera predeterminada). Devuelve la fecha u hora de inicio de la instrucción actual, incluso cuando se encuentra dentro de un bloque de transacciones.

### Sintaxis

```
GETDATE()
```

Los paréntesis son obligatorios.

Tipo de retorno

TIMESTAMP

Ejemplos

En el siguiente ejemplo, se usa la función GETDATE para devolver la marca temporal completa para la fecha actual.

```
select getdate();

timestamp
---------------------
2008-12-04 16:10:43
```

En el siguiente ejemplo, se usa la función GETDATE dentro de la función TRUNC para devolver la fecha actual sin la hora.

```
select trunc(getdate());

trunc
------------
2008-12-04
```

## Función INTERVAL_CMP

INTERVAL_CMP compara dos intervalos y devuelve 1 si el primer intervalo es mayor, -1 si el segundo intervalo es mayor y 0 si los intervalos son iguales. Para obtener más información, consulte Ejemplos de literales de intervalo sin sintaxis de calificador.

Sintaxis

```
INTERVAL_CMP(interval1, interval2)
```

Argumentos

interval1

Un valor literal de intervalo.

interval2

Un valor literal de intervalo.

Tipo de retorno

INTEGER

Ejemplos

En el siguiente ejemplo, se compara el valor de `3 days` con `1 year`.

```
select interval_cmp('3 days','1 year');

interval_cmp
--------------
-1
```

En este ejemplo se compara el valor `7 days` con `1 week`.

```
select interval_cmp('7 days','1 week');

interval_cmp
--------------
0
```

En el siguiente ejemplo, se compara el valor de `1 year` con `3 days`.

```
select interval_cmp('1 year','3 days');

interval_cmp
--------------
1
```

## Función LAST_DAY

LAST_DAY devuelve la fecha del último día del mes que contiene el valor de date (fecha). El tipo de valor devuelto siempre es DATE, sin importar el tipo de datos del argumento date (fecha).

Para obtener más información sobre la recuperación de partes de fecha específicas, consulte Función DATE_TRUNC.

Sintaxis

```
LAST_DAY( { date | timestamp } )
```

Argumentos

date | timestamp

Una columna de tipo de datos DATE o TIMESTAMP, o una expresión que toma implícitamente el valor de un tipo DATE o TIMESTAMP.

Tipo de retorno

DATE

Ejemplos

En el siguiente ejemplo, se devuelve la fecha del último día del mes actual.

```
select last_day(sysdate);

  last_day
------------
 2014-01-31
```

En el siguiente ejemplo, se devuelve la cantidad de tickets vendidos para cada uno de los últimos 7 días del mes. Los valores de la columna SALETIME son marcas temporales.

```
select datediff(day, saletime, last_day(saletime)) as "Days Remaining", sum(qtysold)
from sales
where datediff(day, saletime, last_day(saletime)) < 7
group by 1
order by 1;

days remaining |  sum
---------------+-------
             0 | 10140
             1 | 11187
             2 | 11515
             3 | 11217
             4 | 11446
             5 | 11708
             6 | 10988
```

```
(7 rows)
```

## Función MONTHS_BETWEEN

MONTHS_BETWEEN determina la cantidad de meses que hay entre dos fechas.

Si la primera fecha es posterior a la segunda, el resultado es positivo; de lo contrario, el resultado es negativo.

Si alguno de los argumentos es nulo, el resultado es NULL.

### Sintaxis

```
MONTHS_BETWEEN( date1, date2 )
```

### Argumentos

date1

Una columna de tipo de datos DATE o una expresión que toma implícitamente el valor de un tipo DATE.

date2

Una columna de tipo de datos DATE o una expresión que toma implícitamente el valor de un tipo DATE.

### Tipo de retorno

FLOAT8

La porción de número entero del resultado se basa en la diferencia entre los valores de mes y año de las fechas. La porción de fracción del resultado se calcula a partir de los valores de día y marca temporal de las fechas y supone un mes de 31 días.

Si tanto date1 (fecha1) como date2 (fecha2) tienen el mismo día del mes (por ejemplo, 15/01/14 y 15/02/14) o el último día el mes (por ejemplo, 31/08/14 y 30/09/14), el resultado será un número entero basado en los valores de mes y año de las fechas, sin importar si la porción de marca temporal (en caso de estar presente) coincide.

### Ejemplos

En el siguiente ejemplo, se devuelven los meses entre 18/01/1969 y 18/03/1969.

```
select months_between('1969-01-18', '1969-03-18')
as months;

months
----------
-2
```

En el siguiente ejemplo, se devuelven los meses entre 18/01/1969 y 18/01/1969.

```
select months_between('1969-01-18', '1969-01-18')
as months;

months
----------
0
```

En el siguiente ejemplo, se devuelven los meses entre la primera y la última muestra de un evento.

```
select eventname,
min(starttime) as first_show,
max(starttime) as last_show,
months_between(max(starttime),min(starttime)) as month_diff
from event
group by eventname
order by eventname
limit 5;

eventname         first_show             last_show              month_diff
-------------------------------------------------------------------------
.38 Special       2008-01-21 19:30:00.0  2008-12-25 15:00:00.0  11.12
3 Doors Down      2008-01-03 15:00:00.0  2008-12-01 19:30:00.0  10.94
70s Soul Jam      2008-01-16 19:30:00.0  2008-12-07 14:00:00.0  10.7
A Bronx Tale      2008-01-21 19:00:00.0  2008-12-15 15:00:00.0  10.8
A Catered Affair  2008-01-08 19:30:00.0  2008-12-19 19:00:00.0  11.35
```

## Función NEXT_DAY

NEXT_DAY devuelve la fecha de la primera instancia del día especificado que es posterior a la fecha proporcionada.

Si el valor de day es el mismo día de la semana que la fecha especificada, se devuelve la siguiente instancia de ese día.

### Sintaxis

```
NEXT_DAY( { date | timestamp }, day )
```

### Argumentos

date | timestamp

Una columna de tipo de datos DATE o TIMESTAMP, o una expresión que toma implícitamente el valor de un tipo DATE o TIMESTAMP.

día

Una cadena que tiene el nombre de cualquier día. El uso de mayúsculas no se tiene en cuenta.

Los valores válidos son los siguientes:

| Día | Valores |
|---|---|
| Domingo | Su, Sun, Sunday |
| Lunes | M, Mo, Mon, Monday |
| Martes | Tu, Tue, Tues, Tuesday |
| Miércoles | W, We Wed, Wednesday |
| Jueves | Th, Thu, Thurs, Thursday |
| Viernes | F, Fr, Fri, Friday |
| Sábado | Sa, Sat, Saturday |

### Tipo de retorno

DATE

### Ejemplos

En el siguiente ejemplo, se devuelve la fecha del primer martes después del 20/08/2014.

```
select next_day('2014-08-20','Tuesday');

next_day
-----------
2014-08-26
```

En el siguiente ejemplo se devuelve la fecha del primer martes después del 01/01/2008 a las 5:54:44.

```
select listtime, next_day(listtime, 'Tue') from listing limit 1;

listtime            | next_day
--------------------+-----------
2008-01-01 05:54:44 | 2008-01-08
```

En el siguiente ejemplo, se obtienen las fechas de marketing objetivo para el tercer trimestre.

```
select username, (firstname ||' '|| lastname) as name,
eventname, caldate, next_day (caldate, 'Monday') as marketing_target
from sales, date, users, event
where sales.buyerid = users.userid
and sales.eventid = event.eventid
and event.dateid = date.dateid
and date.qtr = 3
order by marketing_target, eventname, name;

username  |     name          |     eventname        |    caldate   |
 marketing_target
----------+-------------------+----------------------+--------------
+------------------
MBO26QSG  |   Callum Atkinson | .38 Special          |  2008-07-06  | 2008-07-07
WCR50YIU  |   Erasmus Alvarez | A Doll's House       |  2008-07-03  | 2008-07-07
CKT70OIE  |   Hadassah Adkins | Ana Gabriel          |  2008-07-06  | 2008-07-07
VVG07OUO  |   Nathan Abbott   | Armando Manzanero    |  2008-07-04  | 2008-07-07
GEW77SII  |   Scarlet Avila   | August: Osage County |  2008-07-06  | 2008-07-07
ECR71CVS  |   Caryn Adkins    | Ben Folds            |  2008-07-03  | 2008-07-07
KUW82CYU  |   Kaden Aguilar   | Bette Midler         |  2008-07-01  | 2008-07-07
WZE78DJZ  |   Kay Avila       | Bette Midler         |  2008-07-01  | 2008-07-07
HXY04NVE  |   Dante Austin    | Britney Spears       |  2008-07-02  | 2008-07-07
URY81YWF  |   Wilma Anthony   | Britney Spears       |  2008-07-02  | 2008-07-07
```

## Función SYSDATE

SYSDATE devuelve la fecha y hora actual en la zona horaria de la sesión actual (que es UTC de manera predeterminada).

> **Note**
>
> SYSDATE devuelve la fecha y hora de comienzo de la transacción actual, no de la instrucción actual.

### Sintaxis

```
SYSDATE
```

Esta función no requiere argumentos.

### Tipo de retorno

TIMESTAMP

### Ejemplos

En el siguiente ejemplo, se usa la función SYSDATE para devolver la marca temporal completa para la fecha actual.

```
select sysdate;

timestamp
---------------------------
2008-12-04 16:10:43.976353
```

En el siguiente ejemplo, se usa la función SYSDATE dentro de la función TRUNC para devolver la fecha actual sin la hora.

```
select trunc(sysdate);

trunc
------------
```

```
2008-12-04
```

La siguiente consulta devuelve la información de ventas correspondiente a las fechas comprendidas entre la fecha en que se emite la consulta y la fecha que caiga 120 días antes.

```
select salesid, pricepaid, trunc(saletime) as saletime, trunc(sysdate) as now
from sales
where saletime between trunc(sysdate)-120 and trunc(sysdate)
order by saletime asc;

 salesid | pricepaid |  saletime  |    now
---------+-----------+------------+------------
91535    |    670.00 | 2008-08-07 | 2008-12-05
91635    |    365.00 | 2008-08-07 | 2008-12-05
91901    |   1002.00 | 2008-08-07 | 2008-12-05
...
```

## Función TIMEOFDAY

TIMEOFDAY es un alias especial que se usa para devolver el día, la fecha y la hora como un valor de cadena. Devuelve la cadena de hora del día para la instrucción actual, incluso cuando se encuentra dentro de un bloque de transacciones.

Sintaxis

```
TIMEOFDAY()
```

Tipo de retorno

VARCHAR

Ejemplos

En el siguiente ejemplo, se devuelve la fecha y la hora actual mediante el uso de la función TIMEOFDAY.

```
select timeofday();

timeofday
------------
Thu Sep 19 22:53:50.333525 2013 UTC
```

## Función TIMESTAMP_CMP

Compara el valor de dos marcas temporales y devuelve un número entero. Si las marcas temporales son idénticas, la función devuelve `0`. Si la primera marca temporal es mayor, la función devuelve `1`. Si la segunda marca temporal es mayor, la función devuelve `-1`.

### Sintaxis

```
TIMESTAMP_CMP(timestamp1, timestamp2)
```

### Argumentos

timestamp1

Una columna de tipo de datos `TIMESTAMP` o una expresión que toma implícitamente el valor de un tipo `TIMESTAMP`.

timestamp2

Una columna de tipo de datos `TIMESTAMP` o una expresión que toma implícitamente el valor de un tipo `TIMESTAMP`.

### Tipo de retorno

INTEGER

### Ejemplos

En el siguiente ejemplo se comparan las marcas temporales y se muestran los resultados de la comparación.

```
SELECT TIMESTAMP_CMP('2008-01-24 06:43:29', '2008-01-24 06:43:29'),
 TIMESTAMP_CMP('2008-01-24 06:43:29', '2008-02-18 02:36:48'), TIMESTAMP_CMP('2008-02-18
 02:36:48', '2008-01-24 06:43:29');

timestamp_cmp  | timestamp_cmp | timestamp_cmp
---------------+---------------+---------------
             0 |            -1 |             1
```

En el siguiente ejemplo, se compara LISTTIME y SALETIME para una lista. El valor de TIMESTAMP_CMP es `-1` para todos los listados porque la marca temporal para la venta es posterior a la marca temporal para el listado.

```
select listing.listid, listing.listtime,
sales.saletime, timestamp_cmp(listing.listtime, sales.saletime)
from listing, sales
where listing.listid=sales.listid
order by 1, 2, 3, 4
limit 10;

 listid |      listtime       |      saletime       | timestamp_cmp
--------+---------------------+---------------------+---------------
      1 | 2008-01-24 06:43:29 | 2008-02-18 02:36:48 |            -1
      4 | 2008-05-24 01:18:37 | 2008-06-06 05:00:16 |            -1
      5 | 2008-05-17 02:29:11 | 2008-06-06 08:26:17 |            -1
      5 | 2008-05-17 02:29:11 | 2008-06-09 08:38:52 |            -1
      6 | 2008-08-15 02:08:13 | 2008-08-31 09:17:02 |            -1
     10 | 2008-06-17 09:44:54 | 2008-06-26 12:56:06 |            -1
     10 | 2008-06-17 09:44:54 | 2008-07-10 02:12:36 |            -1
     10 | 2008-06-17 09:44:54 | 2008-07-16 11:59:24 |            -1
     10 | 2008-06-17 09:44:54 | 2008-07-22 02:23:17 |            -1
     12 | 2008-07-25 01:45:49 | 2008-08-04 03:06:36 |            -1
(10 rows)
```

En este ejemplo, se muestra que TIMESTAMP_CMP devuelve 0 en caso de marcas temporales idénticas:

```
select listid, timestamp_cmp(listtime, listtime)
from listing
order by 1 , 2
limit 10;

 listid | timestamp_cmp
--------+---------------
      1 |             0
      2 |             0
      3 |             0
      4 |             0
      5 |             0
      6 |             0
      7 |             0
      8 |             0
      9 |             0
     10 |             0
(10 rows)
```

## Función TIMESTAMP_CMP_DATE

TIMESTAMP_CMP_DATE compara el valor de una marca temporal y una fecha. Si los valores de marca temporal y fecha son idénticos, la función devuelve `0`. Si la primera marca temporal es mayor, la función devuelve 1. Si la fecha es mayor, la función devuelve `-1`.

### Sintaxis

```
TIMESTAMP_CMP_DATE(timestamp, date)
```

### Argumentos

timestamp

Una columna de tipo de datos `TIMESTAMP` o una expresión que toma implícitamente el valor de un tipo `TIMESTAMP`.

date

Una columna de tipo de datos `DATE` o una expresión que toma implícitamente el valor de un tipo `DATE`.

### Tipo de retorno

INTEGER

### Ejemplos

En el siguiente ejemplo, se compara LISTTIME con la fecha `2008-06-18`. Las listas hechas después de esa fecha devuelven 1; las hechas antes de ella, `-1`. Los valores LISTTIME son marcas temporales.

```
select listid, listtime,
timestamp_cmp_date(listtime, '2008-06-18')
from listing
order by 1, 2, 3
limit 10;


 listid |      listtime       | timestamp_cmp_date
--------+---------------------+--------------------
      1 | 2008-01-24 06:43:29 |                 -1
```

```
      2 | 2008-03-05 12:25:29 |              -1
      3 | 2008-11-01 07:35:33 |               1
      4 | 2008-05-24 01:18:37 |              -1
      5 | 2008-05-17 02:29:11 |              -1
      6 | 2008-08-15 02:08:13 |               1
      7 | 2008-11-15 09:38:15 |               1
      8 | 2008-11-09 05:07:30 |               1
      9 | 2008-09-09 08:03:36 |               1
     10 | 2008-06-17 09:44:54 |              -1
(10 rows)
```

## Función TIMESTAMP_CMP_TIMESTAMPTZ

TIMESTAMP_CMP_TIMESTAMPTZ compara el valor de una expresión de marca temporal con una expresión de marca temporal con zona horaria. Si los valores de marca temporal y marca temporal con zona horaria son idénticos, la función devuelve `0`. Si la primera marca temporal es mayor, la función devuelve 1. Si la marca temporal con zona horaria es mayor, la función devuelve –1.

### Sintaxis

```
TIMESTAMP_CMP_TIMESTAMPTZ(timestamp, timestamptz)
```

### Argumentos

timestamp

Una columna de tipo de datos `TIMESTAMP` o una expresión que toma implícitamente el valor de un tipo `TIMESTAMP`.

timestamptz

Una columna de tipo de datos `TIMESTAMPTZ` o una expresión que toma implícitamente el valor de un tipo `TIMESTAMPTZ`.

### Tipo de retorno

INTEGER

### Ejemplos

En el siguiente ejemplo se comparan las marcas temporales con las marcas temporales con zona horaria y se muestran los resultados de la comparación.

```
SELECT TIMESTAMP_CMP_TIMESTAMPTZ('2008-01-24 06:43:29', '2008-01-24 06:43:29+00'),
 TIMESTAMP_CMP_TIMESTAMPTZ('2008-01-24 06:43:29', '2008-02-18 02:36:48+00'),
 TIMESTAMP_CMP_TIMESTAMPTZ('2008-02-18 02:36:48', '2008-01-24 06:43:29+00');

timestamp_cmp_timestamptz  | timestamp_cmp_timestamptz | timestamp_cmp_timestamptz
---------------------------+---------------------------+--------------------------
             0             |            -1             |             1
```

## Función TIMESTAMPTZ_CMP

TIMESTAMPTZ_CMP compara el valor de dos valores de marca temporal con zona horaria y devuelve un número entero. Si las marcas temporales son idénticas, la función devuelve `0`. Si la primera marca temporal es mayor cronológicamente, la función devuelve 1. Si la segunda marca temporal es mayor, la función devuelve −1.

### Sintaxis

```
TIMESTAMPTZ_CMP(timestamptz1, timestamptz2)
```

### Argumentos

timestamptz1

Una columna de tipo de datos `TIMESTAMPTZ` o una expresión que toma implícitamente el valor de un tipo `TIMESTAMPTZ`.

timestamptz2

Una columna de tipo de datos `TIMESTAMPTZ` o una expresión que toma implícitamente el valor de un tipo `TIMESTAMPTZ`.

### Tipo de retorno

INTEGER

### Ejemplos

En el siguiente ejemplo se comparan las marcas temporales con zonas horarias y se muestran los resultados de la comparación.

```
SELECT TIMESTAMPTZ_CMP('2008-01-24 06:43:29+00', '2008-01-24 06:43:29+00'),
 TIMESTAMPTZ_CMP('2008-01-24 06:43:29+00', '2008-02-18 02:36:48+00'),
 TIMESTAMPTZ_CMP('2008-02-18 02:36:48+00', '2008-01-24 06:43:29+00');

timestamptz_cmp  | timestamptz_cmp | timestamptz_cmp
-----------------+-----------------+----------------
        0        |       -1        |       1
```

## Función TIMESTAMPTZ_CMP_DATE

TIMESTAMPTZ_CMP_DATE compara el valor de una marca temporal y una fecha. Si los valores de marca temporal y fecha son idénticos, la función devuelve `0`. Si la primera marca temporal es mayor, la función devuelve 1. Si la fecha es mayor, la función devuelve `-1`.

### Sintaxis

```
TIMESTAMPTZ_CMP_DATE(timestamptz, date)
```

### Argumentos

timestamptz

Una columna de tipo de datos `TIMESTAMPTZ` o una expresión que toma implícitamente el valor de un tipo `TIMESTAMPTZ`.

date

Una columna de tipo de datos `DATE` o una expresión que toma implícitamente el valor de un tipo `DATE`.

### Tipo de retorno

INTEGER

### Ejemplos

En el siguiente ejemplo se compara LISTTIME como marca temporal con zona horaria con la fecha `2008-06-18`. Las listas hechas después de esa fecha devuelven `1`; las hechas antes de ella, `-1`.

```
select listid, CAST(listtime as timestamptz) as tstz,
```

```
timestamp_cmp_date(tstz, '2008-06-18')
from listing
order by 1, 2, 3
limit 10;


 listid |          tstz          | timestamptz_cmp_date
--------+------------------------+----------------------
      1 | 2008-01-24 06:43:29+00 |                   -1
      2 | 2008-03-05 12:25:29+00 |                   -1
      3 | 2008-11-01 07:35:33+00 |                    1
      4 | 2008-05-24 01:18:37+00 |                   -1
      5 | 2008-05-17 02:29:11+00 |                   -1
      6 | 2008-08-15 02:08:13+00 |                    1
      7 | 2008-11-15 09:38:15+00 |                    1
      8 | 2008-11-09 05:07:30+00 |                    1
      9 | 2008-09-09 08:03:36+00 |                    1
     10 | 2008-06-17 09:44:54+00 |                   -1
(10 rows)
```

## Función TIMESTAMPTZ_CMP_TIMESTAMP

TIMESTAMPTZ_CMP_TIMESTAMP compara el valor de una expresión de marca temporal con zona horaria con una expresión de marca temporal. Si los valores de marca temporal y marca temporal con zona horaria son idénticos, la función devuelve `0`. Si la marca temporal con zona horaria es mayor cronológicamente, la función devuelve 1. Si la marca temporal es mayor, la función devuelve −1.

### Sintaxis

```
TIMESTAMPTZ_CMP_TIMESTAMP(timestamptz, timestamp)
```

### Argumentos

timestamptz

Una columna de tipo de datos `TIMESTAMPTZ` o una expresión que toma implícitamente el valor de un tipo `TIMESTAMPTZ`.

timestamp

Una columna de tipo de datos `TIMESTAMP` o una expresión que toma implícitamente el valor de un tipo `TIMESTAMP`.

Tipo de retorno

INTEGER

Ejemplos

En el siguiente ejemplo se comparan las marcas temporales con zonas horarias con las marcas temporales y se muestran los resultados de la comparación.

```
SELECT TIMESTAMPTZ_CMP_TIMESTAMP('2008-01-24 06:43:29+00', '2008-01-24 06:43:29'),
 TIMESTAMPTZ_CMP_TIMESTAMP('2008-01-24 06:43:29+00', '2008-02-18 02:36:48'),
 TIMESTAMPTZ_CMP_TIMESTAMP('2008-02-18 02:36:48+00', '2008-01-24 06:43:29');

timestamptz_cmp_timestamp  | timestamptz_cmp_timestamp | timestamptz_cmp_timestamp
---------------------------+---------------------------+--------------------------
             0             |            -1             |             1
```

## Función TIMEZONE

TIMEZONE devuelve una marca temporal para el valor de marca temporal y zona horaria especificados.

Para obtener información y ejemplos acerca de cómo establecer la zona horaria, consulte timezone.

Para obtener información y ejemplos acerca de cómo convertir la zona horaria, consulte CONVERT_TIMEZONE.

Sintaxis

```
TIMEZONE('timezone', { timestamp | timestamptz })
```

Argumentos

timezone

La zona horaria para el valor de retorno. La zona horaria puede especificarse como un nombre (por ejemplo, **'Africa/Kampala'** o **'Singapore'**) o como una abreviatura (por ejemplo, **'UTC'** o **'PDT'**). Para ver una lista de los nombres de zonas horarias compatibles, ejecute el siguiente comando.

```
select pg_timezone_names();
```

Para ver una lista de las abreviaturas de zonas horarias compatibles, ejecute el siguiente comando.

```
select pg_timezone_abbrevs();
```

Para obtener más información y ejemplos, consulta Notas de uso de la zona horaria.

timestamp | timestamptz

Una expresión que da como resultado un valor de tipo TIMESTAMP o TIMESTAMPTZ o un valor que pueda convertirse implícitamente a una marca temporal o a una marca temporal con zona horaria.

Tipo de retorno

TIMESTAMPTZ, cuando se usa con una expresión TIMESTAMP.

TIMESTAMP, cuando se usa con una expresión TIMESTAMPTZ.

Ejemplos

Lo siguiente devuelve una marca temporal para la zona horaria UTC mediante la marca temporal `2008-06-17 09:44:54` de la zona horaria PST.

```
SELECT TIMEZONE('PST', '2008-06-17 09:44:54');

timezone
-----------------------
2008-06-17 17:44:54+00
```

Lo siguiente devuelve una marca temporal para la zona horaria PST mediante la marca temporal con zona horaria UTC `2008-06-17 09:44:54+00`.

```
SELECT TIMEZONE('PST', timestamptz('2008-06-17 09:44:54+00'));

timezone
-----------------------
2008-06-17 01:44:54
```

## Función TO_TIMESTAMP

TO_TIMESTAMP convierte una cadena TIMESTAMP en TIMESTAMPTZ. Para obtener una lista de las funciones de fecha y hora adicionales de Amazon Redshift, consulte Funciones de fecha y hora.

### Sintaxis

```
to_timestamp(timestamp, format)
```

```
to_timestamp (timestamp, format, is_strict)
```

### Argumentos

timestamp

Una cadena que representa un valor de marca temporal en el formato especificado por format. Si este argumento se deja vacío, el valor predeterminado de la marca de tiempo es `0001-01-01 00:00:00`.

format

Un literal de cadena que define el formato del valor de timestamp. Los formatos que incluyen una zona horaria (**`TZ`**, **`tz`** u **`OF`**) no son compatibles como entrada. Para obtener los formatos de marca temporal válidos, consulte Cadenas de formatos de fecha y hora.

is_strict

Un valor booleano opcional que especifica si se devuelve un error si un valor de marca temporal de entrada se encuentra fuera de rango. Cuando is_strict se configura como TRUE, se devuelve un error si hay un valor fuera de rango. Si is_strict se configura como FALSE, que es el valor predeterminado, se aceptan valores de desbordamiento.

### Tipo de retorno

TIMESTAMPTZ

### Ejemplos

En el siguiente ejemplo, se muestra cómo se usa la función TO_TIMESTAMP para convertir una cadena TIMESTAMP en TIMESTAMPTZ.

```
select sysdate, to_timestamp(sysdate, 'YYYY-MM-DD HH24:MI:SS') as second;
```

```
timestamp                  | second
---------------------------   ---------------------
2021-04-05 19:27:53.281812 | 2021-04-05 19:27:53+00
```

Es posible pasar a TO_TIMESTAMP parte de una fecha. Las partes de fecha restantes se establecen a los valores predeterminados. La hora se incluye en el resultado:

```
SELECT TO_TIMESTAMP('2017','YYYY');

to_timestamp
---------------------------
2017-01-01 00:00:00+00
```

La siguiente instrucción SQL convierte la cadena “2011-12-18 24:38:15” en un valor TIMESTAMPTZ. El resultado es un valor TIMESTAMPTZ que se establece al día siguiente porque el número de horas es superior a 24 horas:

```
SELECT TO_TIMESTAMP('2011-12-18 24:38:15', 'YYYY-MM-DD HH24:MI:SS');

to_timestamp
---------------------
2011-12-19 00:38:15+00
```

La siguiente instrucción SQL convierte la cadena “2011-12-18 24:38:15” en un valor TIMESTAMPTZ. El resultado es un error porque el valor de tiempo en la marca de tiempo es superior a 24 horas:

```
SELECT TO_TIMESTAMP('2011-12-18 24:38:15', 'YYYY-MM-DD HH24:MI:SS', TRUE);

ERROR:  date/time field time value out of range: 24:38:15.0
```

## Función TRUNC

Trunca un valor `TIMESTAMP` y devuelve un valor `DATE`.

Esta función también puede truncar un número. Para obtener más información, consulte Función TRUNC.

Sintaxis

```
TRUNC(timestamp)
```

Argumentos

timestamp

Una columna de tipo de datos `TIMESTAMP` o una expresión que toma implícitamente el valor de un tipo `TIMESTAMP`.

Para devolver un valor de marca temporal con `00:00:00` como la hora, convierta el resultado de la función en `TIMESTAMP`.

Tipo de retorno

DATE

Ejemplos

En el siguiente ejemplo, se devuelve la parte de la fecha del resultado de la función SYSDATE (que devuelve una marca temporal).

```
SELECT SYSDATE;

+----------------------------+
|         timestamp          |
+----------------------------+
| 2011-07-21 10:32:38.248109 |
+----------------------------+

SELECT TRUNC(SYSDATE);

+------------+
|   trunc    |
+------------+
| 2011-07-21 |
+------------+
```

En el siguiente ejemplo, se aplica la función TRUNC a una columna `TIMESTAMP`. El tipo de retorno es una fecha.

```
SELECT TRUNC(starttime) FROM event
ORDER BY eventid LIMIT 1;

+------------+
```

```
|   trunc    |
+------------+
| 2008-01-25 |
+------------+
```

En el siguiente ejemplo se devuelve un valor de marca temporal con `00:00:00` como la hora mediante la conversión del resultado de la función TRUNC en `TIMESTAMP`.

```
SELECT CAST((TRUNC(SYSDATE)) AS TIMESTAMP);

+---------------------+
|        trunc        |
+---------------------+
| 2011-07-21 00:00:00 |
+---------------------+
```

## Partes de fecha para funciones de fecha o marca temporal

En la siguiente tabla, se identifican los nombres y las abreviaturas de partes de fecha y de hora que se aceptan como argumentos para las siguientes funciones:

- DATEADD
- DATEDIFF
- DATE_PART
- EXTRACT

| Parte de la fecha o parte de la hora | Abreviaturas |
|---|---|
| milenio, milenios | mil, mils |
| siglo, siglos | c, cent, cents |
| década, décadas | dec, decs |
| tiempo Unix | fecha de inicio (compatible con EXTRACT) |
| año, años | y, yr, yrs |

| Parte de la fecha o parte de la hora | Abreviaturas |
| --- | --- |
| trimestre, trimestres | qtr, qtrs |
| mes, meses | mon, mons |
| semana, semanas | w |
| día de la semana | dayofweek, dow, dw, weekday (compatibles con DATE_PART y Función EXTRACT)<br><br>Devuelve un número entero de 0 a 6, comenzando por domingo.<br><br>**Note**<br>La parte de la fecha DOW se comporta de manera diferente a la parte de fecha (D) que se usa para las cadenas de formato de fecha y hora. D se basa en los números enteros de 1 a 7, donde domingo es 1. Para obtener más información, consulte Cadenas de formatos de fecha y hora. |
| día del año | dayofyear, doy, dy, yearday (compatibles con EXTRACT) |
| día, días | d |
| hora, horas | h, hr, hrs |
| minuto, minutos | m, min, mins |
| segundo, segundos | s, sec, secs |
| milisegundo, milisegundos | ms, msec, msecs, msecond, mseconds, millisec, millisecs, millisecon |
| microsegundo, microsegundos | microsec, microsecs, microsecond, usecond, useconds, us, usec, usecs |

| Parte de la fecha o parte de la hora | Abreviaturas |
| --- | --- |
| zona horaria, timezone_hour, timezone_minute | Compatible solo con EXTRACT para marca temporal con zona horaria (TIMESTAMPTZ). |

Variaciones en resultados con segundos, milisegundos y microsegundos

Cuando diferentes funciones de fechas especifican segundos, milisegundos o microsegundos como partes de fecha, se generan diferencias mínimas en los resultados de las consultas:

- La función EXTRACT devuelve números enteros solo para la parte de fecha especificada e ignora partes de fecha de niveles mayores y menores. Si la parte de fecha especificada es segundos, los milisegundos y los microsegundos no se incluyen en el resultado. Si la parte de fecha especificada es milisegundos, los segundos y los microsegundos no se incluyen. Si la parte de fecha especificada es microsegundos, los segundos y los milisegundos no se incluyen.
- La función DATE_PART devuelve la parte de segundos de la marca temporal completa, sin importar la parte de fecha especificada, por lo que devuelve un valor decimal o un número entero según se requiera.

Por ejemplo, compare los resultados de las siguientes consultas:

```
create table seconds(micro timestamp);

insert into seconds values('2009-09-21 11:10:03.189717');

select extract(sec from micro) from seconds;

date_part
-----------
3

select date_part(sec, micro) from seconds;

pgdate_part
-------------
3.189717
```

### Notas acerca de CENTURY, EPOCH, DECADE y MIL

#### CENTURY o CENTURIES

Amazon Redshift interpreta que CENTURY comienza en el año ###1 y termina en el año ###0:

```
select extract (century from timestamp '2000-12-16 12:21:13');
date_part
-----------
20

select extract (century from timestamp '2001-12-16 12:21:13');
date_part
-----------
21
```

#### EPOCH

La implementación de EPOCH en Amazon Redshift está asociada a 1970-01-01 00:00:00.000000, sin importar la zona horaria donde resida el clúster. Podría ser necesario desplazar los resultados de la diferencia en horas según la zona horaria donde se encuentre el clúster.

En el siguiente ejemplo, se muestra lo siguiente:

1. Crea una tabla denominada EVENT_EXAMPLE basada en la tabla EVENT. Este comando CREATE AS usa la función DATE_PART para crear una columna de fecha (denominada PGDATE_PART de manera predeterminada) para almacenar el valor de fecha de inicio de cada evento.
2. Selecciona la columna y el tipo de dato de EVENT_EXAMPLE de PG_TABLE_DEF.
3. Selecciona EVENTNAME, STARTTIME y PGDATE_PART de la tabla EVENT_EXAMPLE para visualizar los diferentes formatos de fecha y hora.
4. Selecciona EVENTNAME y STARTTIME de EVENT EXAMPLE como están. Convierte los valores de fecha de inicio en PGDATE_PART mediante el uso de un intervalo de 1 segundo a una marca temporal sin zona horaria y devuelve los resultados en una columna denominada CONVERTED_TIMESTAMP.

```
create table event_example
as select eventname, starttime, date_part(epoch, starttime) from event;

select "column", type from pg_table_def where tablename='event_example';
```

```
     column    |            type
---------------+-----------------------------
 eventname     | character varying(200)
 starttime     | timestamp without time zone
 pgdate_part   | double precision
(3 rows)
```

```
select eventname, starttime, pgdate_part from event_example;

   eventname          |      starttime      | pgdate_part
----------------------+---------------------+-------------
 Mamma Mia!           | 2008-01-01 20:00:00 |  1199217600
 Spring Awakening     | 2008-01-01 15:00:00 |  1199199600
 Nas                  | 2008-01-01 14:30:00 |  1199197800
 Hannah Montana       | 2008-01-01 19:30:00 |  1199215800
 K.D. Lang            | 2008-01-01 15:00:00 |  1199199600
 Spamalot             | 2008-01-02 20:00:00 |  1199304000
 Macbeth              | 2008-01-02 15:00:00 |  1199286000
 The Cherry Orchard   | 2008-01-02 14:30:00 |  1199284200
 Macbeth              | 2008-01-02 19:30:00 |  1199302200
 Demi Lovato          | 2008-01-02 19:30:00 |  1199302200


select eventname,
starttime,
timestamp with time zone 'epoch' + pgdate_part * interval '1 second' AS
 converted_timestamp
from event_example;

       eventname      |      starttime      | converted_timestamp
----------------------+---------------------+---------------------
 Mamma Mia!           | 2008-01-01 20:00:00 | 2008-01-01 20:00:00
 Spring Awakening     | 2008-01-01 15:00:00 | 2008-01-01 15:00:00
 Nas                  | 2008-01-01 14:30:00 | 2008-01-01 14:30:00
 Hannah Montana       | 2008-01-01 19:30:00 | 2008-01-01 19:30:00
 K.D. Lang            | 2008-01-01 15:00:00 | 2008-01-01 15:00:00
 Spamalot             | 2008-01-02 20:00:00 | 2008-01-02 20:00:00
 Macbeth              | 2008-01-02 15:00:00 | 2008-01-02 15:00:00
 The Cherry Orchard   | 2008-01-02 14:30:00 | 2008-01-02 14:30:00
 Macbeth              | 2008-01-02 19:30:00 | 2008-01-02 19:30:00
 Demi Lovato          | 2008-01-02 19:30:00 | 2008-01-02 19:30:00
 ...
```

DECADE o DECADES

Amazon Redshift interpreta DECADE o DECADES DATEPART en función del calendario común. Por ejemplo, debido a que el calendario común comienza a partir del año 1, la primera década (década 1) es de 0001-01-01 a 0009-12-31 y la segunda década (década 2) es de 0010-01-01 a 0019-12-31. Por ejemplo, la década 201 se extiende de 01/01/2001 a 31/12/2009:

```
select extract(decade from timestamp '1999-02-16 20:38:40');
date_part
-----------
200

select extract(decade from timestamp '2000-02-16 20:38:40');
date_part
-----------
201

select extract(decade from timestamp '2010-02-16 20:38:40');
date_part
-----------
202
```

MIL o MILS

Amazon Redshift interpreta que MIL comienza el primer día del año #001 y termina el último día del año #000:

```
select extract (mil from timestamp '2000-12-16 12:21:13');
date_part
-----------
2

select extract (mil from timestamp '2001-12-16 12:21:13');
date_part
-----------
3
```

# Funciones hash

Temas

- Función CHECKSUM

- Función farmFingerprint64
- Función FUNC_SHA1
- Función FNV_HASH
- Función MD5
- Función SHA
- Función SHA1
- Función SHA2
- MURMUR3_32_HASH

Una función hash es una función matemática que convierte un valor de entrada numérico en otro valor.

## Función CHECKSUM

Computa un valor de suma de comprobación para elaborar un índice hash.

Sintaxis

```
CHECKSUM(expression)
```

Argumento

expresión

La expresión de entrada debe ser un tipo de datos VARCHAR, INTEGER o DECIMAL.

Tipo de retorno

La función CHECKSUM devuelve un valor entero.

Ejemplo

En el siguiente ejemplo, se computa un valor de suma de comprobación para la columna COMMISSION:

```
select checksum(commission)
from sales
order by salesid
limit 10;
```

```
checksum
----------
10920
1140
5250
2625
2310
5910
11820
2955
8865
975
(10 rows)
```

## Función farmFingerprint64

Calcula el valor de farmhash del argumento de entrada mediante la función `Fingerprint64`.

### Sintaxis

```
farmFingerprint64(expression)
```

### Argumento

expresión

La expresión de entrada debe ser un tipo de datos `VARCHAR` o `VARBYTE`.

### Tipo de retorno

La función `farmFingerprint64` devuelve `BIGINT`.

### Ejemplo

En el siguiente ejemplo, se devuelve el valor `farmFingerprint64` de `Amazon Redshift` que es una entrada como tipo de datos `VARCHAR`.

```
SELECT farmFingerprint64('Amazon Redshift');
```

```
  farmfingerprint64
```

```
---------------------
 8085098817162212970
```

En el siguiente ejemplo, se devuelve el valor `farmFingerprint64` de `Amazon Redshift` que es una entrada como tipo de datos `VARBYTE`.

```
SELECT farmFingerprint64('Amazon Redshift'::varbyte);
```

```
  farmfingerprint64
---------------------
 8085098817162212970
```

## Función FUNC_SHA1

Sinónimo de la función SHA1.

Consulte Función SHA1.

## Función FNV_HASH

Calcula la función hash no criptográfica FNV-1a de 64 bits para todos los tipos de datos básicos.

Sintaxis

```
FNV_HASH(value [, seed])
```

Argumentos

value

El valor de entrada al que se le aplicará un algoritmo hash. Amazon Redshift utiliza la representación binaria del valor para aplicar un algoritmo hash al valor de entrada; por ejemplo, a los valores INTEGER se les aplica un algoritmo hash con 4 bytes y a los valores BIGINT se les aplica un algoritmo hash con 8 bytes. Además, las entradas de hash CHAR y VARCHAR no ignoran los espacios finales.

valor de inicialización

El valor de inicialización BIGINT de la función hash es opcional. Si no se proporciona, Amazon Redshift utiliza el valor de inicialización FNV predeterminado. Esto permite combinar el hash de varias columnas sin conversiones ni concatenaciones.

Tipo de retorno

BIGINT

Ejemplo

En los siguientes ejemplos se devuelve el hash FNV de un número, la cadena "Amazon Redshift" y la concatenación de los dos.

```
select fnv_hash(1);
        fnv_hash
----------------------
 -5968735742475085980
(1 row)
```

```
select fnv_hash('Amazon Redshift');
      fnv_hash
--------------------
 7783490368944507294
(1 row)
```

```
select fnv_hash('Amazon Redshift', fnv_hash(1));
       fnv_hash
----------------------
 -2202602717770968555
(1 row)
```

Notas de uso

- Para calcular el hash de una tabla con varias columnas, puede calcular el hash FNV de la primera columna y pasarlo como valor de inicialización al hash de la segunda columna. A continuación, pasa el hash FNV de la segunda columna como valor de inicialización al hash de la tercera columna.

  En el siguiente ejemplo, se crean valores de inicialización para aplicar un algoritmo hash a una tabla con varias columnas.

  ```
  select fnv_hash(column_3, fnv_hash(column_2, fnv_hash(column_1))) from sample_table;
  ```

- La misma propiedad se puede utilizar para calcular el hash de una concatenación de cadenas.

```
select fnv_hash('abcd');
       fnv_hash
---------------------
 -281581062704388899
(1 row)
```

```
select fnv_hash('cd', fnv_hash('ab'));
      fnv_hash
---------------------
 -281581062704388899
(1 row)
```

- La función hash utiliza el tipo de entrada para determinar el número de bytes para aplicar un algoritmo hash. Utilice la conversión para aplicar un tipo específico, si es necesario.

  En los ejemplos siguientes se utilizan diferentes tipos de entrada para producir resultados diferentes.

```
select fnv_hash(1::smallint);
      fnv_hash
--------------------
 589727492704079044
(1 row)
```

```
select fnv_hash(1);
       fnv_hash
----------------------
 -5968735742475085980
(1 row)
```

```
select fnv_hash(1::bigint);
       fnv_hash
----------------------
 -8517097267634966620
(1 row)
```

## Función MD5

Utiliza la función hash criptográfica MD5 para convertir una cadena de longitud variable en una cadena de 32 caracteres que es una representación textual del valor hexadecimal de una suma de comprobación de 128 bits.

### Sintaxis

```
MD5(string)
```

### Argumentos

string

Una cadena de longitud variable.

### Tipo de retorno

La función MD5 devuelve una cadena de 32 caracteres que es una representación textual del valor hexadecimal de una suma de comprobación de 128 bits.

### Ejemplos

En el siguiente ejemplo, se muestra el valor de 128 bits para la cadena 'Amazon Redshift':

```
select md5('Amazon Redshift');
md5
----------------------------------
f7415e33f972c03abd4f3fed36748f7a
(1 row)
```

## Función SHA

Sinónimo de la función SHA1.

Consulte Función SHA1.

## Función SHA1

La función SHA1 usa la función criptográfica hash SHA1 para convertir una cadena de longitud variable en una cadena de 40 caracteres que es una representación textual del valor hexadecimal de una suma de comprobación de 160 bits.

### Sintaxis

SHA1 es sinónimo de Función SHA y Función FUNC_SHA1.

```
SHA1(string)
```

### Argumentos

string

Una cadena de longitud variable.

### Tipo de retorno

La función SHA1 devuelve una cadena de 40 caracteres que es una representación textual del valor hexadecimal de una suma de comprobación de 160 bits.

### Ejemplo

El siguiente ejemplo devuelve el valor de 160 bits para la palabra “Amazon Redshift”:

```
select sha1('Amazon Redshift');
```

## Función SHA2

La función SHA2 utiliza la función hash criptográfica SHA2 para convertir una cadena de longitud variable en una cadena de caracteres. La cadena de caracteres es una representación de texto del valor hexadecimal de la suma de comprobación con el número especificado de bits.

### Sintaxis

```
SHA2(string, bits)
```

### Argumentos

string

Una cadena de longitud variable.

integer

El número de bits en las funciones hash. Los valores válidos son 0 (igual que 256), 224, 256, 384 y 512.

Tipo de retorno

La función SHA2 devuelve una cadena de caracteres que es una representación textual del valor hexadecimal de la suma de comprobación o una cadena vacía si el número de bits no es válido.

Ejemplo

El siguiente ejemplo devuelve el valor de 256 bits para la palabra “Amazon Redshift”:

```
select sha2('Amazon Redshift', 256);
```

## MURMUR3_32_HASH

La función MURMUR3_32_HASH calcula el hash no criptográfico Murmur3A de 32 bits para todos los tipos de datos comunes, incluidos los tipos numéricos y de cadena.

Sintaxis

```
MURMUR3_32_HASH(value [, seed])
```

Argumentos

value

El valor de entrada al que se aplica hash. Amazon Redshift realiza el hash de la representación binaria del valor de entrada. Este comportamiento es similar al de Función FNV_HASH, pero el valor se convierte a la representación binaria indicada por la especificación hash Murmur3 de 32 bits de Apache Iceberg.

valor de inicialización

El valor de inicialización INT de la función hash. Este argumento es opcional. Si no se proporciona, Amazon Redshift utiliza el valor de inicialización predeterminado de 0. Esto permite combinar el hash de varias columnas sin conversiones ni concatenaciones.

Tipo de retorno

La función devuelve un valor INT.

Ejemplo

En los siguientes ejemplos se devuelve el hash Murmur3 de un número, la cadena “Amazon Redshift” y la concatenación de los dos.

```
select MURMUR3_32_HASH(1);

    MURMUR3_32_HASH
----------------------
 1392991556
(1 row)
```

```
select MURMUR3_32_HASH('Amazon Redshift');

    MURMUR3_32_HASH
----------------------
 -1563580564
(1 row)
```

```
select MURMUR3_32_HASH('Amazon Redshift', MURMUR3_32_HASH(1));

    MURMUR3_32_HASH
----------------------
 -1346554171
(1 row)
```

Notas de uso

Para calcular el hash de una tabla con varias columnas, puede calcular el hash Murmur3 de la primera columna y pasarlo como valor de inicialización al hash de la segunda columna. A continuación, pasa el hash Murmur3 de la segunda columna como valor de inicialización al hash de la tercera columna.

En el siguiente ejemplo, se crean valores de inicialización para aplicar un algoritmo hash a una tabla con varias columnas.

```
select MURMUR3_32_HASH(column_3, MURMUR3_32_HASH(column_2, MURMUR3_32_HASH(column_1)))
 from sample_table;
```

La misma propiedad se puede utilizar para calcular el hash de una concatenación de cadenas.

```
select MURMUR3_32_HASH('abcd');

   MURMUR3_32_HASH
----------------------
 1139631978
(1 row)
```

```
select MURMUR3_32_HASH('cd', MURMUR3_32_HASH('ab'));

   MURMUR3_32_HASH
----------------------
 1711522338
(1 row)
```

La función hash utiliza el tipo de entrada para determinar el número de bytes para aplicar un algoritmo hash. Utilice la conversión para aplicar un tipo específico, si es necesario.

En los ejemplos siguientes se utilizan diferentes tipos de entrada para producir resultados diferentes.

```
select MURMUR3_32_HASH(1, MURMUR3_32_HASH(1));

   MURMUR3_32_HASH
--------------------
 -1193428387
(1 row)
```

```
select MURMUR3_32_HASH(1);

   MURMUR3_32_HASH
----------------------
 1392991556
(1 row)
```

```
select MURMUR3_32_HASH(1, MURMUR3_32_HASH(2));
```

```
   MURMUR3_32_HASH
----------------------
 1179621905
(1 row)
```

# Funciones HyperLogLog

A continuación, encontrará la descripción de las funciones HyperLogLog para SQL que son compatibles con Amazon Redshift.

Temas

- Funciones HLL
- Función HLL_CREATE_SKETCH
- Función HLL_CARDINALITY
- Función HLL_COMBINE
- Función HLL_COMBINE_SKETCHES

## Funciones HLL

La función HLL devuelve la cardinalidad de HyperLogLog de los valores de la expresión de entrada. La función HLL es compatible con cualquier tipo de datos, excepto HLLSKETCH. La función HLL ignora los valores NULL. Cuando no hay filas en una tabla o todas las filas son NULL, la cardinalidad resultante es 0.

Sintaxis

```
HLL (aggregate_expression)
```

Argumento

expresión_de_agregación

Toda expresión válida que proporcione el valor a una agrupación, como el nombre de una columna. Esta función admite cualquier tipo de datos como entrada, excepto HLLSKETCH, GEOMETRY, GEOGRAPHY y VARBYTE.

Tipo de retorno

La función HLL devuelve un valor BIGINT o INT8.

Ejemplos

El siguiente ejemplo devuelve la cardinalidad de la columna `an_int` en la tabla `a_table`.

```
CREATE TABLE a_table(an_int INT);
INSERT INTO a_table VALUES (1), (2), (3), (4);

SELECT hll(an_int) AS cardinality FROM a_table;
cardinality
-------------
4
```

## Función HLL_CREATE_SKETCH

La función HLL_CREATE_SKETCH devuelve un tipo de datos HLLSKETCH que encapsula los valores de expresión de entrada. La función HLL_CREATE_SKETCH es compatible con cualquier tipo de datos e ignora los valores NULL. Cuando no hay filas en una tabla o todas las filas son NULL, el boceto resultante no tiene pares índice-valor como `{"version":1,"logm":15,"sparse":{"indices":[],"values":[]}}`.

Sintaxis

```
HLL_CREATE_SKETCH (aggregate_expression)
```

Argumento

expresión_de_agregación

Toda expresión válida que proporcione el valor a una agrupación, como el nombre de una columna. Los valores NULL se omiten. Esta función admite cualquier tipo de datos como entrada, excepto HLLSKETCH, GEOMETRY, GEOGRAPHY y VARBYTE.

Tipo de retorno

La función HLL_CREATE_SKETCH devuelve un valor HLLSKETCH.

### Ejemplos

El siguiente ejemplo devuelve el tipo HLLSKETCH para la columna `an_int` en la tabla `a_table`. Se utiliza un objeto JSON para representar un boceto de HyperLogLog disperso al momento de importar, exportar o imprimir los bocetos. Se utiliza la representación de una cadena (en formato Base64) para representar un boceto de HyperLogLog denso.

```
CREATE TABLE a_table(an_int INT);
INSERT INTO a_table VALUES (1), (2), (3), (4);

SELECT hll_create_sketch(an_int) AS sketch FROM a_table;
sketch
--------------------------------------------------------------------------------------------------
{"version":1,"logm":15,"sparse":{"indices":
[20812342,20850007,22362299,47158030],"values":[1,2,1,1]}}
(1 row)
```

## Función HLL_CARDINALITY

La función HLL_CARDINALITY devuelve la cardinalidad del tipo de datos HLLSKETCH de entrada.

### Sintaxis

```
HLL_CARDINALITY (hllsketch_expression)
```

### Argumento

hllsketch_expression

Una expresión válida que toma el valor de un tipo HLLSKETCH, como el nombre de una columna. El valor de entrada es el tipo de datos HLLSKETCH.

### Tipo de retorno

La función HLL_CARDINALITY devuelve un valor BIGINT o INT8.

### Ejemplos

El siguiente ejemplo devuelve la cardinalidad de la columna `sketch` en la tabla `hll_table`.

```
CREATE TABLE a_table(an_int INT, b_int INT);
```

```
INSERT INTO a_table VALUES (1,1), (2,1), (3,1), (4,1), (1,2), (2,2), (3,2), (4,2),
 (5,2), (6,2);

CREATE TABLE hll_table (sketch HLLSKETCH);
INSERT INTO hll_table select hll_create_sketch(an_int) from a_table group by b_int;

SELECT hll_cardinality(sketch) AS cardinality FROM hll_table;
cardinality
-------------
6
4
(2 rows)
```

## Función HLL_COMBINE

La función de agrupación HLL_COMBINE devuelve un tipo de datos HLLSKETCH que combina todos los valores HLLSKETCH de entrada.

La combinación de dos o más bocetos de HyperLogLog es un nuevo HLLSKETCH que encapsula información sobre la unión de los distintos valores que representa cada boceto de entrada. Después de combinar bocetos, Amazon Redshift extrae la cardinalidad de la unión de dos o más conjuntos de datos. Para obtener más información acerca de cómo combinar varios bocetos, consulte Ejemplo: devolver un boceto de HyperLogLog a partir de la combinación de varios bocetos.

### Sintaxis

```
HLL_COMBINE (hllsketch_expression)
```

### Argumento

hllsketch_expression

Una expresión válida que toma el valor de un tipo HLLSKETCH, como el nombre de una columna. El valor de entrada es el tipo de datos HLLSKETCH.

### Tipo de retorno

La función HLL_COMBINE devuelve un tipo HLLSKETCH.

### Ejemplos

El siguiente ejemplo devuelve los valores combinados de HLLSKETCH en la tabla `hll_table`.

```
CREATE TABLE a_table(an_int INT, b_int INT);
INSERT INTO a_table VALUES (1,1), (2,1), (3,1), (4,1), (1,2), (2,2), (3,2), (4,2),
 (5,2), (6,2);

CREATE TABLE hll_table (sketch HLLSKETCH);
INSERT INTO hll_table select hll_create_sketch(an_int) from a_table group by b_int;

SELECT hll_combine(sketch) AS sketches FROM hll_table;
sketches
--------------------------------------------------------------------------------------------
{"version":1,"logm":15,"sparse":{"indices":
[20812342,20850007,22362299,40314817,42650774,47158030],"values":[1,2,1,3,2,1]}}
(1 row)
```

## Función HLL_COMBINE_SKETCHES

HLL_COMBINE_SKETCHES es una función escalar que toma como entrada dos valores de HLLSKETCH y los combina en un solo HLLSKETCH.

La combinación de dos o más bocetos de HyperLogLog es un nuevo HLLSKETCH que encapsula información sobre la unión de los distintos valores que representa cada boceto de entrada.

### Sintaxis

```
HLL_COMBINE_SKETCHES (hllsketch_expression1, hllsketch_expression2)
```

### Argumento

hllsketch_expression1 y hllsketch_expression2

Una expresión válida que toma el valor de un tipo HLLSKETCH, como el nombre de una columna.

### Tipo de retorno

La función HLL_COMBINE_SKETCHES devuelve un tipo HLLSKETCH.

### Ejemplos

El siguiente ejemplo devuelve los valores combinados de HLLSKETCH en la tabla `hll_table`.

```
WITH tbl1(x, y)
     AS (SELECT Hll_create_sketch(1),
```

```
                Hll_create_sketch(2)
        UNION ALL
        SELECT Hll_create_sketch(3),
                Hll_create_sketch(4)
        UNION ALL
        SELECT Hll_create_sketch(5),
                Hll_create_sketch(6)
        UNION ALL
        SELECT Hll_create_sketch(7),
                Hll_create_sketch(8)),
    tbl2(x, y)
    AS (SELECT Hll_create_sketch(9),
                Hll_create_sketch(10)
        UNION ALL
        SELECT Hll_create_sketch(11),
                Hll_create_sketch(12)
        UNION ALL
        SELECT Hll_create_sketch(13),
                Hll_create_sketch(14)
        UNION ALL
        SELECT Hll_create_sketch(15),
                Hll_create_sketch(16)
        UNION ALL
        SELECT Hll_create_sketch(NULL),
                Hll_create_sketch(NULL)),
    tbl3(x, y)
    AS (SELECT *
        FROM   tbl1
        UNION ALL
        SELECT *
        FROM   tbl2)
SELECT Hll_combine_sketches(x, y)
FROM   tbl3;
```

# Funciones JSON

## Temas

- Función JSON_PARSE
- Función CAN_JSON_PARSE
- Función JSON_SERIALIZE
- Función JSON_SERIALIZE_TO_VARBYTE

- Funciones JSON basadas en texto

> **Note**
>
> Le recomendamos que utilice las siguientes funciones para trabajar con JSON:
>
> - Función JSON_PARSE
> - Función CAN_JSON_PARSE
> - Función JSON_SERIALIZE
> - Función JSON_SERIALIZE_TO_VARBYTE
>
> Con JSON_PARSE, solo necesita convertir el texto JSON en un valor de tipo SUPER una vez en la ingesta, después de lo cual puede operar con los valores SUPER. Amazon Redshift analiza los valores SUPER de forma más eficiente que VARCHAR, que es el resultado de las funciones JSON basadas en texto. Para obtener más información acerca de cómo trabajar con el tipo de datos SUPER, vaya a Datos semiestructurados en Amazon Redshift.

Cuando necesita almacenar un conjunto relativamente pequeño de pares clave-valor, puede ahorrar espacio al almacenar los datos en formato JSON. Debido a que las cadenas JSON se pueden almacenar en una única columna, utilizar JSON puede ser más eficiente que almacenar los datos en formato de tabla. Por ejemplo, suponga que tiene una tabla dispersa, donde necesita tener muchas columnas para representar completamente todos los atributos posibles, pero la mayoría de los valores son NULL para cualquier fila o columna dada. Al usar JSON para almacenamiento, puede almacenar los datos para una fila en pares clave-valor en una única cadena JSON y eliminar las columnas de tabla poco pobladas.

Además, puede modificar fácilmente las cadenas JSON para almacenar pares clave-valor adicionales cuando el esquema JSON cambia sin necesidad de agregar columnas a una tabla.

Recomendamos utilizar JSON con moderación. JSON no es una buena opción para almacenar grandes conjuntos de datos porque, al almacenar datos dispersos en una única columna, JSON no utiliza la arquitectura de almacén de columnas de Amazon Redshift. Aunque Amazon Redshift admite las funciones JSON en las columnas CHAR y VARCHAR, recomendamos utilizar SUPER para procesar los datos con el formato de serialización de JSON. SUPER utiliza una representación sin esquema posterior al análisis con la que se pueden consultar datos jerárquicos de manera eficiente.

Para obtener más información acerca del tipo de datos SUPER, consulte Datos semiestructurados en Amazon Redshift.

JSON utiliza cadenas de texto con cifrado UTF-8, por lo que las cadenas JSON se pueden almacenar como tipos de datos CHAR o VARCHAR.

Las cadenas JSON deben tener el formato JSON adecuado, conforme a las siguientes reglas:

- El JSON a nivel raíz puede ser un objeto JSON o una matriz JSON. Un objeto JSON es un conjunto no ordenado de pares clave-valor separados por comas y delimitado con llaves.

  Por ejemplo, `{"one":1, "two":2}`
- Una matriz JSON es un conjunto ordenado de valores separados por comas delimitado entre corchetes.

  A continuación se muestra un ejemplo: `["first", {"one":1}, "second", 3, null]`
- Las matrices JSON utilizan un índice basado en cero; el primer elemento en una matriz está en la posición 0. En un par clave:valor de JSON, la clave es una cadena con comillas dobles.
- El valor JSON puede ser cualquiera de los siguientes valores:
  - Objeto JSON
  - array
  - cadena
    - Se representa con comillas dobles
  - número
    - Incluye números enteros, decimales y flotantes
  - booleano
  - null
- Los objetos y las matrices vacíos son valores JSON válidos.
- Los campos JSON distinguen entre mayúsculas y minúsculas.
- Se ignoran los espacios en blanco entre los elementos estructurales de JSON (como `{ }, [ ]`).

Las funciones JSON de Amazon Redshift y el comando COPY de Amazon Redshift utilizan los mismos métodos para trabajar con datos con formato JSON. Para obtener más información acerca de cómo trabajar con JSON;, consulte COPY de formato JSON

## Función JSON_PARSE

La función JSON_PARSE analiza los datos con formato JSON y los convierte en la representación SUPER.

Para capturar el tipo de datos SUPER mediante el comando INSERT o UPDATE, utilice la función JSON_PARSE. Cuando utiliza JSON_PARSE() para analizar cadenas JSON en valores SUPER, se aplican determinadas restricciones. Para obtener información adicional, consulta Opciones de análisis de SUPER.

### Sintaxis

```
JSON_PARSE( {json_string | binary_value} )
```

### Argumentos

json_string

Una expresión que devuelve JSON serializado como tipo VARBYTE o VARCHAR.

binary_value

Valor binario de tipo VARBYTE.

### Tipo de retorno

SUPER

### Ejemplos

Para convertir la matriz JSON `[10001,10002,"abc"]` en el tipo de datos SUPER, utilice el siguiente ejemplo.

```
SELECT JSON_PARSE('[10001,10002,"abc"]');

+---------------------+
|     json_parse      |
+---------------------+
| [10001,10002,"abc"] |
+---------------------+
```

Para asegurarse de que la función convirtió la matriz JSON en el tipo de datos SUPER, utilice el siguiente ejemplo. Para obtener más información, consulte Función JSON_TYPEOF

```
SELECT JSON_TYPEOF(JSON_PARSE('[10001,10002,"abc"]'));

+-------------+
| json_typeof |
+-------------+
| array       |
+-------------+
```

## Función CAN_JSON_PARSE

La función CAN_JSON_PARSE analiza los datos en formato JSON y devuelve `true` si el resultado se puede convertir en un valor SUPER mediante la función JSON_PARSE.

Sintaxis

```
CAN_JSON_PARSE( {json_string | binary_value} )
```

Argumentos

json_string

Una expresión que devuelve objetos JSON serializados en el formato VARCHAR.

binary_value

Valor binario de tipo VARBYTE.

Tipo de retorno

BOOLEAN

Notas de uso

- CAN_JSON_PARSE devuelve el valor false para cadenas vacías. Devuelve NULL si el argumento de entrada es nulo.

### Ejemplos

En el siguiente ejemplo se muestra CAN_JSON_PARSE en ejecución en una matriz JSON correctamente formada mediante una condición CASE. Devuelve el valor true, por lo que Amazon Redshift ejecuta la función JSON_PARSE en el valor de ejemplo.

```
SELECT CASE
            WHEN CAN_JSON_PARSE('[10001,10002,"abc"]')
            THEN JSON_PARSE('[10001,10002,"abc"]')
        END;

 case
---------------------
'[10001,10002,"abc"]'
```

En el siguiente ejemplo se muestra CAN_JSON_PARSE en ejecución en un valor que no tiene formato JSON mediante una condición CASE. Devuelve el valor false, por lo que Amazon Redshift devuelve en su lugar el segmento de la cláusula ELSE de la condición CASE.

```
SELECT CASE
            WHEN CAN_JSON_PARSE('This is a string.')
            THEN JSON_PARSE('This is a string.')
            ELSE 'This is not JSON.'
        END;

 case
---------------------
"This is not JSON."
```

## Función JSON_SERIALIZE

La función JSON_SERIALIZE serializa una expresión SUPER en una representación JSON textual en función de RFC 8259. Para obtener más información acerca de dicho RFC, consulte Formato de intercambio de datos de notación de objetos de JavaScript (JSON).

El límite de tamaño de SUPER es aproximadamente el mismo que el límite de bloque y el límite de VARCHAR es menor que el límite de tamaño de SUPER. Por lo tanto, la función JSON_SERIALIZE devuelve un error cuando el formato JSON excede el límite de VARCHAR del sistema. Para comprobar el tamaño de una expresión SUPER, consulte la función JSON_SIZE.

### Sintaxis

```
JSON_SERIALIZE(super_expression)
```

### Argumentos

super_expression

Una expresión o columna SUPER.

### Tipo de retorno

VARCHAR

> **Note**
>
> El valor VARCHAR devuelto es siempre una cadena JSON no nula. Si super_expression es NULL, JSON_SERIALIZE devuelve la cadena JSON `'null'`.

### Ejemplos

Para serializar un valor de SUPER en una cadena, use el siguiente ejemplo.

```
SELECT JSON_SERIALIZE(JSON_PARSE('[10001,10002,"abc"]'));

+---------------------+
|   json_serialize    |
+---------------------+
| [10001,10002,"abc"] |
+---------------------+
```

## Función JSON_SERIALIZE_TO_VARBYTE

La función JSON_SERIALIZE_TO_VARBYTE convierte un valor SUPER en una cadena JSON similar a JSON_SERIALIZE(), pero se almacena en un valor VARBYTE en su lugar.

### Sintaxis

```
JSON_SERIALIZE_TO_VARBYTE(super_expression)
```

Argumentos

super_expression

Una expresión o columna SUPER.

Tipo de retorno

VARBYTE

Ejemplos

Para serializar un valor SUPER y devolver el resultado en formato VARBYTE, use el siguiente ejemplo.

```
SELECT JSON_SERIALIZE_TO_VARBYTE(JSON_PARSE('[10001,10002,"abc"]'));

+----------------------------------------+
|       json_serialize_to_varbyte        |
+----------------------------------------+
| 5b31303030312c31303030322c22616263225d |
+----------------------------------------+
```

Para serializar un valor SUPER y convertir el resultado en formato VARCHAR, use el siguiente ejemplo. Para obtener más información, consulte Función CAST.

```
SELECT CAST((JSON_SERIALIZE_TO_VARBYTE(JSON_PARSE('[10001,10002,"abc"]'))) AS VARCHAR);

+---------------------------+
| json_serialize_to_varbyte |
+---------------------------+
| [10001,10002,"abc"]       |
+---------------------------+
```

## Funciones JSON basadas en texto

Las siguientes funciones analizan los valores JSON como VARCHAR. Para analizar JSON, le recomendamos que utilice las siguientes funciones, que analizan los valores JSON como SUPER. Amazon Redshift analiza los valores SUPER de forma más eficiente que VARCHAR.

- Función JSON_PARSE

- Función CAN_JSON_PARSE
- Función JSON_SERIALIZE
- Función JSON_SERIALIZE_TO_VARBYTE

Temas

- Función IS_VALID_JSON
- Función IS_VALID_JSON_ARRAY
- Función JSON_ARRAY_LENGTH
- Función JSON_EXTRACT_ARRAY_ELEMENT_TEXT
- Función JSON_EXTRACT_PATH_TEXT

## Función IS_VALID_JSON

> **Note**
>
> CAN_JSON_PARSE y sus funciones asociadas analizan los valores JSON como SUPER, que Amazon Redshift analiza de forma más eficiente que VARCHAR.
> En lugar de usar IS_VALID_JSON, le recomendamos que valide las cadenas JSON con Función CAN_JSON_PARSE.

La función IS_VALID_JSON valida una cadena JSON. La función devuelve un valor booleano de `true` si la cadena tiene un formato JSON correcto o `false` si la cadena no tiene el formato correcto. Para validar una matriz JSON, use Función IS_VALID_JSON_ARRAY

Para obtener más información, consulte Funciones JSON.

### Sintaxis

```
IS_VALID_JSON('json_string')
```

### Argumentos

json_string

Cadena o expresión que se evalúa como una cadena JSON.

Tipo de retorno

BOOLEAN

Ejemplos

Para crear una tabla e insertar cadenas JSON para su comprobación, use el siguiente ejemplo.

```
CREATE TABLE test_json(id int IDENTITY(0,1), json_strings VARCHAR);

-- Insert valid JSON strings --
INSERT INTO test_json(json_strings) VALUES
('{"a":2}'),
('{"a":{"b":{"c":1}}}'),
('{"a": [1,2,"b"]}');

-- Insert invalid JSON strings --
INSERT INTO test_json(json_strings) VALUES
('{{}}'),
('{1:"a"}'),
('[1,2,3]');
```

Para validar las cadenas del ejemplo anterior, use el siguiente ejemplo.

```
SELECT id, json_strings, IS_VALID_JSON(json_strings)
FROM test_json
ORDER BY id;

+----+--------------------+---------------+
| id |    json_strings    | is_valid_json |
+----+--------------------+---------------+
|  0 | {"a":2}            | true          |
|  4 | {"a":{"b":{"c":1}}} | true          |
|  8 | {"a": [1,2,"b"]}   | true          |
| 12 | {{}}               | false         |
| 16 | {1:"a"}            | false         |
| 20 | [1,2,3]            | false         |
+----+--------------------+---------------+
```

Función IS_VALID_JSON_ARRAY

> **Note**
>
> JSON_PARSE y sus funciones asociadas analizan los valores JSON como SUPER, que Amazon Redshift analiza de forma más eficiente que VARCHAR.
> En lugar de usar IS_VALID_JSON_ARRAY, le recomendamos que analice las cadenas JSON con Función JSON_PARSE para obtener un valor SUPER. A continuación, utilice la función Función IS_ARRAY para confirmar que la matriz está formada correctamente.

La función IS_VALID_JSON_ARRAY valida una matriz JSON. La función devuelve un valor booleano de `true` si la matriz tiene un formato JSON correcto o `false` si la matriz no tiene el formato correcto. Para validar una cadena JSON, use Función IS_VALID_JSON

Para obtener más información, consulte Funciones JSON.

Sintaxis

```
IS_VALID_JSON_ARRAY('json_array')
```

Argumentos

json_array

Cadena o expresión que se evalúa como una matriz JSON.

Tipo de retorno

BOOLEAN

Ejemplos

Para crear una tabla e insertar cadenas JSON para su comprobación, use el siguiente ejemplo.

```
CREATE TABLE test_json_arrays(id int IDENTITY(0,1), json_arrays VARCHAR);

-- Insert valid JSON array strings --
INSERT INTO test_json_arrays(json_arrays)
VALUES('[]'),
```

```
('["a","b"]'),
('["a",["b",1,["c",2,3,null]]]');

-- Insert invalid JSON array strings --
INSERT INTO test_json_arrays(json_arrays)
VALUES('{"a":1}'),
('a'),
('[1,2,]');
```

Para validar las cadenas del ejemplo anterior, use el siguiente ejemplo.

```
SELECT json_arrays, IS_VALID_JSON_ARRAY(json_arrays)
FROM test_json_arrays ORDER BY id;

+------------------------------+---------------------+
|         json_arrays          | is_valid_json_array |
+------------------------------+---------------------+
| []                           | true                |
| ["a","b"]                    | true                |
| ["a",["b",1,["c",2,3,null]]] | true                |
| {"a":1}                      | false               |
| a                            | false               |
| [1,2,]                       | false               |
+------------------------------+---------------------+
```

Función JSON_ARRAY_LENGTH

> **Note**
>
> JSON_PARSE y sus funciones asociadas analizan los valores JSON como SUPER, que Amazon Redshift analiza de forma más eficiente que VARCHAR.
> En lugar de usar JSON_ARRAY_LENGTH, le recomendamos que analice las cadenas JSON con Función JSON_PARSE para obtener un valor SUPER. A continuación, use Función get_array_length para obtener la longitud de la matriz.

La función JSON_ARRAY_LENGTH devuelve la cantidad de elementos en la matriz exterior de una cadena JSON. Si el argumento null_if_invalid está establecido en `true` y la cadena JSON no es válida, la función devuelve NULL en lugar de un error.

Para obtener más información, consulte Funciones JSON.

### Sintaxis

```
JSON_ARRAY_LENGTH('json_array' [, null_if_invalid ] )
```

### Argumentos

json_array

Una matriz JSON con formato adecuado.

null_if_invalid

(Opcional) Un valor BOOLEAN que especifica que se devuelva NULL si la cadena JSON de entrada no es válida en lugar de devolver un error. Para devolver NULL si la cadena JSON no es válida, especifique `true` (`t`). Para devolver un error si la cadena JSON no es válida, especifique `false` (`f`). El valor predeterminado es `false`.

### Tipo de retorno

INTEGER

### Ejemplos

Para devolver la cantidad de elementos en la matriz, use el siguiente ejemplo.

```
SELECT JSON_ARRAY_LENGTH('[11,12,13,{"f1":21,"f2":[25,26]},14]');

+-------------------+
| json_array_length |
+-------------------+
|                 5 |
+-------------------+
```

Para devolver un error porque la cadena JSON no es válida, use el siguiente ejemplo.

```
SELECT JSON_ARRAY_LENGTH('[11,12,13,{"f1":21,"f2":[25,26]},14');

ERROR: invalid json array object [11,12,13,{"f1":21,"f2":[25,26]},14
```

Para establecer null_if_invalid en true, para que la instrucción devuelva NULL en lugar de devolver un error para indicar que la cadena JSON no es válida, use el siguiente ejemplo.

```
SELECT JSON_ARRAY_LENGTH('[11,12,13,{"f1":21,"f2":[25,26]},14',true);

+------------------+
| json_array_length |
+------------------+
| NULL              |
+------------------+
```

Función JSON_EXTRACT_ARRAY_ELEMENT_TEXT

> **Note**
>
> JSON_PARSE y sus funciones asociadas analizan los valores JSON como SUPER, que Amazon Redshift analiza de forma más eficiente que VARCHAR.
> En lugar de usar JSON_EXTRACT_ARRAY_ELEMENT_TEXT, le recomendamos que analice las cadenas JSON con Función JSON_PARSE para obtener un valor SUPER. A continuación, consulte el elemento que desee utilizando su índice de matriz, con la sintaxis de `value[element position]`. Para obtener más información sobre cómo consultar los elementos de una matriz en valores SUPER, vaya a Consulta de datos semiestructurados.

La función JSON_EXTRACT_ARRAY_ELEMENT_TEXT devuelve un elemento de la matriz JSON en la matriz extrema de una cadena JSON utilizando un índice basado en cero. El primer elemento en una matriz está en posición 0. Si el índice es negativo o está fuera de los límites, JSON_EXTRACT_ARRAY_ELEMENT_TEXT devuelve NULL. Si el argumento null_if_invalid está establecido en TRUE y la cadena JSON no es válida, la función devuelve NULL en lugar de un error.

Para obtener más información, consulte Funciones JSON.

Sintaxis

```
JSON_EXTRACT_ARRAY_ELEMENT_TEXT('json string', pos [, null_if_invalid ] )
```

Argumentos

json_string

Una cadena JSON con formato adecuado.

pos

Un valor INTEGER que representa el índice del elemento de matriz que se devolverá, utilizando un índice de matriz basado en cero.

null_if_invalid

(Opcional) Un valor BOOLEAN que especifica que se devuelva NULL si la cadena JSON de entrada no es válida en lugar de devolver un error. Para devolver NULL si la cadena JSON no es válida, especifique `true` (`t`). Para devolver un error si la cadena JSON no es válida, especifique `false` (`f`). El valor predeterminado es `false`.

Tipo de retorno

VARCHAR

Una cadena VARCHAR que representa el elemento de matriz JSON al que se hace referencia en pos.

Ejemplos

Para devolver el elemento de matriz en la posición 2, que es el tercer elemento de un índice de matriz basado en cero, use el siguiente ejemplo.

```
SELECT JSON_EXTRACT_ARRAY_ELEMENT_TEXT('[111,112,113]', 2);

+---------------------------------+
| json_extract_array_element_text |
+---------------------------------+
|                             113 |
+---------------------------------+
```

Para devolver un error porque la cadena JSON no es válida, use el siguiente ejemplo.

```
SELECT JSON_EXTRACT_ARRAY_ELEMENT_TEXT('["a",["b",1,["c",2,3,null,]]]',1);

ERROR: invalid json array object ["a",["b",1,["c",2,3,null,]]]
```

Para establecer null_if_invalid en true, para que la instrucción devuelva NULL en lugar de devolver un error para indicar que la cadena JSON no es válida, use el siguiente ejemplo.

```
SELECT JSON_EXTRACT_ARRAY_ELEMENT_TEXT('["a",["b",1,["c",2,3,null,]]]',1,true);

+---------------------------------+
| json_extract_array_element_text |
+---------------------------------+
| NULL                            |
+---------------------------------+
```

Considere las siguientes instrucciones de ejemplo. Si la cadena JSON proporcionada o el índice son NULL, JSON_EXTRACT_ARRAY_ELEMENT_TEXT devuelve NULL independientemente del valor de cualquier otro parámetro.

```
--Statement where json_string is NULL.
SELECT json_extract_array_element_text(NULL, 0)

 json_extract_array_element_text
---------------------------------
                            NULL

--Statement where pos is NULL and json_string is invalid JSON.
SELECT json_extract_array_element_text('invalid_json', NULL);

 json_extract_array_element_text
---------------------------------
                            NULL

--Statement where json_string is NULL and null_if_invalid is FALSE.
SELECT json_extract_array_element_text(NULL, 0, FALSE);

 json_extract_array_element_text
---------------------------------
                            NULL
```

Considere las siguientes instrucciones de ejemplo. Cuando null_if_invalid es TRUE, JSON_EXTRACT_ARRAY_ELEMENT_TEXT devuelve NULL cuando json_string es JSON no válido. Si null_if_invalid es FALSE o no se ha establecido, la función devuelve un error cuando json_string no es válido.

```
--Statement with invalid JSON where null_if_invalid is TRUE.
SELECT json_extract_array_element_text('invalid_json', 0, TRUE);

 json_extract_array_element_text
---------------------------------
                            NULL

--Statement with invalid JSON where null_if_invalid is FALSE.
SELECT json_extract_array_element_text('invalid_json', 0);

ERROR:  JSON parsing error
```

Considere el siguiente ejemplo, donde json_string es JSON válido y pos hace referencia a un valor `null` JSON. En este caso, JSON_EXTRACT_ARRAY_ELEMENT_TEXT devuelve NULL, independientemente del valor de null_if_invalid.

```
--Statement selecting a null value.
SELECT json_extract_array_element_text('[null]', 0);

  json_extract_array_element_text
---------------------------------
                            NULL
```

Función JSON_EXTRACT_PATH_TEXT

> **Note**
>
> JSON_PARSE y sus funciones asociadas analizan los valores JSON como SUPER, que Amazon Redshift analiza de forma más eficiente que VARCHAR.
> En lugar de usar JSON_EXTRACT_PATH_TEXT, le recomendamos que analice las cadenas JSON con Función JSON_PARSE para obtener un valor SUPER. A continuación, consulte el elemento que desee utilizando la sintaxis de `value.attribute`. Para obtener más información sobre cómo consultar los elementos de una matriz en valores SUPER, vaya a Consulta de datos semiestructurados.

La función JSON_EXTRACT_PATH_TEXT devuelve el valor del par clave-valor al que se hace referencia en una serie de elementos de ruta de una cadena JSON. La ruta JSON se puede anidar hasta un máximo de cinco niveles de profundidad. Los elementos de ruta

distinguen entre mayúsculas y minúsculas. Si un elemento de ruta no existe en la cadena JSON, JSON_EXTRACT_PATH_TEXT devuelve NULL.

Si el argumento null_if_invalid está establecido en TRUE y la cadena JSON no es válida, la función devuelve NULL en lugar de un error.

JSON_EXTRACT_PATH_TEXT tiene un máximo de tamaño de datos de 64 KB. Por lo tanto, si algún registro JSON tiene más de 64 KB, al procesarlo con JSON_EXTRACT_PATH_TEXT se produce un error.

Para obtener información sobre funciones JSON adicionales, consulte Funciones JSON. Para obtener más información acerca de cómo trabajar con JSON, consulte COPY de formato JSON.

Sintaxis

```
JSON_EXTRACT_PATH_TEXT('json_string', 'path_elem' [,'path_elem'[, …] ]
 [, null_if_invalid ] )
```

Argumentos

json_string

Una cadena JSON con formato adecuado.

path_elem

Un elemento de ruta en una cadena JSON. Se necesita al menos un elemento de ruta. Se pueden especificar elementos de ruta adicionales, hasta un máximo de cinco niveles de profundidad.

null_if_invalid

(Opcional) Un valor BOOLEAN que especifica que se devuelva NULL si la cadena JSON de entrada no es válida en lugar de devolver un error. Para devolver NULL si la cadena JSON no es válida, especifique TRUE (t). Para devolver un error si la cadena JSON no es válida, especifique FALSE (f). El valor predeterminado es FALSE.

En una cadena JSON, Amazon Redshift reconoce a \n como un carácter de línea nueva y a \t como un carácter de tabulación. Para cargar una barra inversa, aplique escape con una barra inversa (\\). Para obtener más información, consulte Caracteres de escape en JSON.

Tipo de retorno

VARCHAR

Una cadena VARCHAR que representa el valor JSON al que se hace referencia en los elementos de ruta.

Ejemplos

Para devolver el valor para la ruta `'f4', 'f6'`, use el siguiente ejemplo.

```
SELECT JSON_EXTRACT_PATH_TEXT('{"f2":{"f3":1},"f4":{"f5":99,"f6":"star"}}','f4', 'f6');

+------------------------+
| json_extract_path_text |
+------------------------+
| star                   |
+------------------------+
```

Para devolver un error porque la cadena JSON no es válida, use el siguiente ejemplo.

```
SELECT JSON_EXTRACT_PATH_TEXT('{"f2":{"f3":1},"f4":{"f5":99,"f6":"star"}','f4', 'f6');

ERROR: invalid json object {"f2":{"f3":1},"f4":{"f5":99,"f6":"star"}
```

Para establecer null_if_invalid en TRUE, para que la instrucción devuelva NULL por cadena JSON no válida en lugar de devolver un error, use el siguiente ejemplo.

```
SELECT JSON_EXTRACT_PATH_TEXT('{"f2":{"f3":1},"f4":{"f5":99,"f6":"star"}','f4',
 'f6',true);

+------------------------+
| json_extract_path_text |
+------------------------+
| NULL                   |
+------------------------+
```

Considere el siguiente ejemplo, que selecciona el valor de la ruta `'farm', 'barn', 'color'`, donde el valor recuperado se encuentra en el tercer nivel. Utilice el siguiente ejemplo. Este ejemplo está formateado con una herramienta de lint JSON para facilitar su lectura.

```
SELECT JSON_EXTRACT_PATH_TEXT('{
    "farm": {
        "barn": {
            "color": "red",
            "feed stocked": true
        }
    }
}', 'farm', 'barn', 'color');
+------------------------+
| json_extract_path_text |
+------------------------+
| red                    |
+------------------------+
```

Para devolver NULL porque falta el elemento `'color'`, use el siguiente ejemplo. Este ejemplo está formateado con una herramienta de lint JSON.

```
SELECT JSON_EXTRACT_PATH_TEXT('{
    "farm": {
        "barn": {}
    }
}', 'farm', 'barn', 'color');

+------------------------+
| json_extract_path_text |
+------------------------+
| NULL                   |
+------------------------+
```

Si el JSON es válido, el intento de extracción de un elemento que falta devuelve NULL.

Para devolver el valor para la ruta `'house', 'appliances', 'washing machine', 'brand'`, use el siguiente ejemplo.

```
SELECT JSON_EXTRACT_PATH_TEXT('{
  "house": {
    "address": {
      "street": "123 Any St.",
      "city": "Any Town",
      "state": "FL",
      "zip": "32830"
    },
```

```
    "bathroom": {
      "color": "green",
      "shower": true
    },
    "appliances": {
      "washing machine": {
        "brand": "Any Brand",
        "color": "beige"
      },
      "dryer": {
        "brand": "Any Brand",
        "color": "white"
      }
    }
  }
}', 'house', 'appliances', 'washing machine', 'brand');

+------------------------+
| json_extract_path_text |
+------------------------+
| Any Brand              |
+------------------------+
```

En el siguiente ejemplo, se crea una tabla de ejemplo y se rellena con valores SUPER. A continuación, se devuelve el valor de la ruta `'f2'` de ambas filas.

```
CREATE TABLE json_example(id INT, json_text SUPER);

INSERT INTO json_example VALUES
(1, JSON_PARSE('{"f2":{"f3":1},"f4":{"f5":99,"f6":"star"}}')),
(2, JSON_PARSE('{
    "farm": {
        "barn": {
            "color": "red",
            "feed stocked": true
        }
    }
}'));

SELECT * FROM json_example;
id          | json_text
------------+-------------------------------------------
1           | {"f2":{"f3":1},"f4":{"f5":99,"f6":"star"}}
```

```
2          | {"farm":{"barn":{"color":"red","feed stocked":true}}}


SELECT id, JSON_EXTRACT_PATH_TEXT(JSON_SERIALIZE(json_text), 'f2') FROM json_example;

id         | json_text
-----------+--------------------------------------------
1          | {"f3":1}
2          |
```

Considere las siguientes instrucciones de ejemplo. El valor de path_elem proporcionado es NULL, por lo que JSON_EXTRACT_PATH_TEXT devuelve NULL independientemente del valor de cualquier otro parámetro.

```
--Statement where path_elem is NULL and json_string is valid JSON.
SELECT JSON_EXTRACT_PATH_TEXT('{"f2":{"f3":1},"f4":{"f5":99,"f6":"star"}}',NULL);

 json_extract_path_text
------------------------
                  NULL

--Statement where only one path_elem is NULL.
SELECT JSON_EXTRACT_PATH_TEXT('{"f2":{"f3":1},"f4":{"f5":99,"f6":"star"}}','f4',NULL);

 json_extract_path_text
------------------------
                  NULL

--Statement where path_elem is NULL and json_string is invalid JSON.
SELECT json_extract_path_text('invalid_json', NULL);

 json_extract_path_text
------------------------
                  NULL

--Statement where path_elem is NULL and null_if_invalid is FALSE.
SELECT json_extract_path_text(NULL, 0, FALSE);

 json_extract_path_text
------------------------
                  NULL
```

Considere las siguientes instrucciones de ejemplo. Cuando null_if_invalid es TRUE, JSON_EXTRACT_PATH_TEXT devuelve NULL cuando json_string es JSON no válido. Si null_if_invalid es FALSE o no se ha establecido, la función devuelve un error cuando json_string no es válido.

```
--Statement with invalid JSON where null_if_invalid is TRUE.
SELECT json_extract_path_text('invalid_json', 0, TRUE);

 json_extract_path_text
------------------------
                  NULL

--Statement with invalid JSON where null_if_invalid is FALSE.
SELECT json_extract_path_text('invalid_json', 0, FALSE);

ERROR:  JSON parsing error
```

Considere los siguientes ejemplos, donde json_string es JSON válido y path_elem hace referencia a un valor `null` JSON. En este caso, JSON_EXTRACT_PATH_TEXT devuelve NULL. Del mismo modo, cuando path_elem hace referencia a un valor inexistente, JSON_EXTRACT_PATH_TEXT devuelve NULL, independientemente del valor de null_if_invalid.

```
--Statement selecting a null value.
SELECT json_extract_path_text('[null]', 0);

  json_extract_path_text
-------------------------
                   NULL

--Statement selecting a non-existing value.
SELECT json_extract_path_text('{}', 'a');

  json_extract_path_text
-------------------------
                   NULL
```

## Funciones de machine learning.

A través del uso de machine learning (ML) de Amazon Redshift ML, puede formar modelos de ML con instrucciones SQL e invocarlos en consultas SQL para generar predicciones. Con la explicabilidad del modelo de Amazon Redshift incorpora valores de importancia de características

para ayudar a comprender cómo contribuye cada atributo de los datos de formación al resultado previsto.

A continuación, encontrará la descripción de las funciones de machine learning para SQL que son compatibles con Amazon Redshift.

Temas


- Función EXPLAIN_MODEL


## Función EXPLAIN_MODEL

La función EXPLAIN_MODEL devuelve un tipo de datos SUPER que contiene un informe de explicabilidad del modelo en formato JSON. El informe de explicabilidad contiene información sobre el valor de Shapley para todas las características del modelo.

Actualmente, la función EXPLAIN_MODEL solo es compatible con los modelos XGBoost AUTO ON o AUTO OFF.

Cuando el informe de explicabilidad no está disponible, la función devuelve estados que muestran el progreso del modelo. Entre ellos se encuentran `Waiting for training job to complete`, `Waiting for processing job to complete` y `Processing job failed`.

Cuando se ejecuta la instrucción CREATE MODEL, el estado de la explicación pasa a ser `Waiting for training job to complete`. Cuando se ha formado el modelo y se envía una solicitud de explicación, el estado de esta pasa a ser `Waiting for processing job to complete`. Cuando la explicación del modelo se completa correctamente, queda a disposición el informe de explicabilidad completo. De lo contrario, el estado pasa a ser `Processing job failed`.

Cuando ejecute la instrucción CREATE MODEL, puede utilizar el parámetro `MAX_RUNTIME` opcional para especificar la cantidad máxima de tiempo que debe durar el entrenamiento. Una vez que la creación del modelo alcanza esa duración, Amazon Redshift deja de crear el modelo. Si alcanza ese límite de tiempo mientras crea un modelo de piloto automático, Amazon Redshift le devolverá el mejor modelo hasta el momento. La explicabilidad del modelo está disponible una vez que finaliza el entrenamiento del modelo, por lo que si `MAX_RUNTIME` se establece a una duración baja, es posible que el informe de explicabilidad no esté disponible. El tiempo de entrenamiento varía y depende de la complejidad del modelo, el tamaño de los datos y otros factores.

### Sintaxis

```
EXPLAIN_MODEL ('schema_name.model_name')
```

### Argumento

schema_name

El nombre del esquema. Si no se especifica schema_name, se selecciona el esquema actual.

model_name

El nombre del modelo. El nombre del modelo en un esquema debe ser único.

### Tipo de retorno

La función EXPLAIN_MODEL devuelve un tipo de datos SUPER, tal y como se muestra a continuación.

```
{"version":"1.0","explanations":{"kernel_shap":{"label0":{"global_shap_values":
{"x0":0.05,"x1":0.10,"x2":0.30,"x3":0.15},"expected_value":0.50}}}}
```

### Ejemplos

El siguiente ejemplo devuelve el estado de la explicación `waiting for training job to complete`.

```
select explain_model('customer_churn_auto_model');
                 explain_model
-------------------------------------------------------
{"explanations":"waiting for training job to complete"}
(1 row)
```

Cuando la explicación del modelo se completa correctamente, queda a disposición el informe de explicabilidad completo de la siguiente manera.

```
select explain_model('customer_churn_auto_model');
                                        explain_model
-------------------------------------------------------------------------------------------------
{"version":"1.0","explanations":{"kernel_shap":{"label0":{"global_shap_values":
{"x0":0.05386043365892927,"x1":0.10801289723274592,"x2":0.23227865827017378,"x3":0.067668513394
```

```
(1 row)
```

Dado que la función EXPLAIN_MODEL devuelve el tipo de datos SUPER, es posible consultar el informe de explicabilidad. De este modo, se pueden extraer valores de Shapley de `global_shap_values`, `expected_value` o de características específicas.

En el siguiente ejemplo, se extrae `global_shap_values` para el modelo.

```
select json_table.report.explanations.kernel_shap.label0.global_shap_values from
 (select explain_model('customer_churn_auto_model') as report) as json_table;
                                                      global_shap_values
-------------------------------------------------------------------------------------------------------
{"state":0.10983770427197151,"account_length":0.1772441398408543,"area_code":0.0862682396863959
(1 row)
```

En el siguiente ejemplo, se extrae `global_shap_values` para la característica x0.

```
select json_table.report.explanations.kernel_shap.label0.global_shap_values.x0 from
 (select explain_model('customer_churn_auto_model') as report) as json_table;
          x0
------------------------
  0.05386043365892927
(1 row)
```

Si el modelo se crea en un esquema específico y se tiene acceso al modelo creado, entonces es posible realizar una consulta sobre la explicación del modelo, tal y como se muestra a continuación.

```
-- Check the current schema
SHOW search_path;
   search_path
-----------------
  $user, public
(1 row)
-- If you have the privilege to access the model explanation
-- in `test_schema`
SELECT explain_model('test_schema.test_model_name');
                      explain_model
-----------------------------------------------------------
{"explanations":"waiting for training job to complete"}
(1 row)
```

# Funciones matemáticas

Temas


- Símbolos de operadores matemáticos
- Función ABS
- Función ACOS
- Función ASIN
- Función ATAN
- Función ATAN2
- Función CBRT
- Función CEILING (o CEIL)
- Función COS
- Función COT
- Función DEGREES
- Función DEXP
- Función DLOG1
- Función DLOG10
- Función EXP
- Función FLOOR
- Función LN
- Función LOG
- Función MOD
- Función PI
- Función POWER
- Función RADIANS
- Función RANDOM
- Función ROUND
- Función SIN
- Función SIGN

- Función SQRT
- Función TAN
- Función TRUNC

En esta sección, se describen los operadores y las funciones matemáticas compatibles con Amazon Redshift.

## Símbolos de operadores matemáticos

En la tabla siguiente, se muestran los operadores matemáticos admitidos.

Operadores admitidos

| Operador | Descripción | Ejemplo | Resultado |
|---|---|---|---|
| + | suma | 2 + 3 | 5 |
| - | resta | 2 - 3 | -1 |
| * | multiplicación | 2 * 3 | 6 |
| / | división | 4/2 | 2 |
| % | módulo | 5 % 4 | 1 |
| ^ | potencia | 2,0 ^ 3,0 | 8 |
| \|/ | raíz cuadrada | \| / 25,0 | 5 |
| \|\|/ | raíz cúbica | \|\| / 27,0 | 3 |
| @ | valor absoluto | @ -5,0 | 5 |
| << | cambio bit a bit hacia la izquierda | 1 << 4 | 16 |

| Operador | Descripción | Ejemplo | Resultado |
|---|---|---|---|
| >> | cambio bit a bit hacia la derecha | 8 >> 2 | 2 |
| & | bit a bit y | 8 & 2 | 0 |

Ejemplos

En los siguientes ejemplos se utiliza la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para calcular la comisión pagada más un 2.00 $ para una transacción dada, utilice el siguiente ejemplo.

```
SELECT
    commission,
    (commission + 2.00) AS comm
FROM
    sales
WHERE
    salesid = 10000;

+------------+-------+
| commission | comm  |
+------------+-------+
|      28.05 | 30.05 |
+------------+-------+
```

Para calcular el 20 % del precio de venta para una transacción dada, utilice el siguiente ejemplo.

```
SELECT pricepaid, (pricepaid * .20) as twentypct
FROM sales
WHERE salesid=10000;

+-----------+-----------+
| pricepaid | twentypct |
+-----------+-----------+
|       187 |      37.4 |
```

```
+-----------+-----------+
```

Para prever las ventas de tickets según un patrón de crecimiento continuo, use el siguiente ejemplo. En este ejemplo, la subconsulta devuelve la cantidad de tickets vendidos en 2008. El resultado se multiplica exponencialmente por un índice de crecimiento continuo del 5% durante 10 años.

```
SELECT (SELECT SUM(qtysold) FROM sales, date
WHERE sales.dateid=date.dateid AND year=2008)^((5::float/100)*10) AS qty10years;

+------------------+
|    qty10years    |
+------------------+
| 587.664019657491 |
+------------------+
```

Para encontrar el precio total pagado y la comisión por ventas con un ID de fecha que sea mayor que o igual a 2000, use el siguiente ejemplo. Luego, se resta la comisión total del precio total pagado.

```
SELECT SUM(pricepaid) AS sum_price, dateid,
SUM(commission) AS sum_comm, (SUM(pricepaid) - SUM(commission)) AS value
FROM sales
WHERE dateid >= 2000
GROUP BY dateid
ORDER BY dateid
LIMIT 10;

+-----------+--------+----------+-----------+
| sum_price | dateid | sum_comm |   value   |
+-----------+--------+----------+-----------+
|    305885 |   2000 | 45882.75 | 260002.25 |
|    316037 |   2001 | 47405.55 | 268631.45 |
|    358571 |   2002 | 53785.65 | 304785.35 |
|    366033 |   2003 | 54904.95 | 311128.05 |
|    307592 |   2004 |  46138.8 |  261453.2 |
|    333484 |   2005 |  50022.6 |  283461.4 |
|    317670 |   2006 |  47650.5 |  270019.5 |
|    351031 |   2007 | 52654.65 | 298376.35 |
|    313359 |   2008 | 47003.85 | 266355.15 |
|    323675 |   2009 | 48551.25 | 275123.75 |
+-----------+--------+----------+-----------+
```

## Función ABS

ABS calcula el valor absoluto de un número, donde ese número puede ser un valor literal o una expresión que tome el valor de un número.

### Sintaxis

```
ABS(number)
```

### Argumentos

número

Número o expresión que toma el valor de un número. Puede ser el tipo SMALLINT, INTEGER, BIGINT, DECIMAL, FLOAT4, FLOAT8 o SUPER.

### Tipo de retorno

ABS devuelve el mismo tipo de datos como su argumento.

### Ejemplos

Para calcular el valor absoluto de -38, utilice el ejemplo siguiente.

```
SELECT ABS(-38);

+-----+
| abs |
+-----+
|  38 |
+-----+
```

Para calcular el valor absoluto de (14-76), utilice el ejemplo siguiente.

```
SELECT ABS(14-76);

+-----+
| abs |
+-----+
|  62 |
```

```
+-----+
```

## Función ACOS

ACOS es una función trigonométrica que devuelve el arcocoseno de un número. El valor de retorno está en radianes y se encuentra entre `0` y `PI`.

### Sintaxis

```
ACOS(number)
```

### Argumentos

número

El parámetro de entrada es un número de `DOUBLE PRECISION`.

### Tipo de retorno

`DOUBLE PRECISION`

### Ejemplos

Para devolver el arcoseno de `-1`, use el siguiente ejemplo.

```
SELECT ACOS(-1);

+-------------------+
|       acos        |
+-------------------+
| 3.141592653589793 |
+-------------------+
```

Para convertir el arcocoseno de `.5` a un número equivalente de grados, use el siguiente ejemplo.

```
SELECT (ACOS(.5) * 180/(SELECT PI())) AS degrees;

+-------------------+
|      degrees      |
```

```
+-------------------+
| 60.00000000000001 |
+-------------------+
```

## Función ASIN

ASIN es una función trigonométrica que devuelve el arcoseno de un número. El valor de retorno está en radianes y se encuentra entre `PI/2` y `-PI/2`.

### Sintaxis

```
ASIN(number)
```

### Argumentos

número

El parámetro de entrada es un número de `DOUBLE PRECISION`.

### Tipo de retorno

`DOUBLE PRECISION`

### Ejemplos

Para devolver el arcoseno de 1, use el siguiente ejemplo.

```
SELECT ASIN(1) AS halfpi;

+--------------------+
|       halfpi       |
+--------------------+
| 1.5707963267948966 |
+--------------------+
```

Para convertir el arcoseno de `.5` a un número equivalente de grados, use el siguiente ejemplo.

```
SELECT (ASIN(.5) * 180/(SELECT PI())) AS degrees;

+--------------------+
```

```
|      degrees       |
+--------------------+
| 30.000000000000004 |
+--------------------+
```

## Función ATAN

ATAN es una función trigonométrica que devuelve la arcotangente de un número. El valor de retorno está en radianes y se encuentra entre `-PI` y `PI`.

### Sintaxis

```
ATAN(number)
```

### Argumentos

número

El parámetro de entrada es un número de `DOUBLE PRECISION`.

### Tipo de retorno

`DOUBLE PRECISION`

### Ejemplos

Para devolver la arcotangente de 1 y multiplicarla por 4, use el siguiente ejemplo.

```
SELECT ATAN(1) * 4 AS pi;

+-------------------+
|        pi         |
+-------------------+
| 3.141592653589793 |
+-------------------+
```

Para convertir la arcotangente de 1 a un número equivalente de grados, use el siguiente ejemplo.

```
SELECT (ATAN(1) * 180/(SELECT PI())) AS degrees;
```

```
+---------+
| degrees |
+---------+
|      45 |
+---------+
```

## Función ATAN2

ATAN2 es una función trigonométrica que devuelve la arcotangente de un número dividido por otro número. El valor de retorno está en radianes y se encuentra entre `PI/2` y `-PI/2`.

### Sintaxis

```
ATAN2(number1, number2)
```

### Argumentos

number1

Un número de `DOUBLE PRECISION`.

number2

Un número de `DOUBLE PRECISION`.

### Tipo de retorno

`DOUBLE PRECISION`

### Ejemplos

Para devolver la arcotangente de `2/2` y multiplicarla por `4`, use el siguiente ejemplo.

```
SELECT ATAN2(2,2) * 4 AS PI;

+-------------------+
|        pi         |
+-------------------+
| 3.141592653589793 |
+-------------------+
```

Para convertir la arcotangente de `1/0` (que evalúa a 0) al número equivalente de grados, use el siguiente ejemplo.

```
SELECT (ATAN2(1,0) * 180/(SELECT PI())) AS degrees;

+---------+
| degrees |
+---------+
|      90 |
+---------+
```

## Función CBRT

La función CBRT es una función matemática que calcula la raíz cúbica de un número dado.

### Sintaxis

```
CBRT(number)
```

### Argumentos

CBRT toma un número de `DOUBLE PRECISION` como argumento.

### Tipo de retorno

`DOUBLE PRECISION`

### Ejemplos

El siguiente ejemplo utiliza la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para calcular la raíz cúbica de la comisión pagada para una transacción dada, utilice el siguiente ejemplo.

```
SELECT CBRT(commission) FROM sales WHERE salesid=10000;

+--------------------+
|        cbrt        |
+--------------------+
| 3.0383953904884344 |
```

```
+--------------------+
```

## Función CEILING (o CEIL)

La función CEILING o CEIL se usa para redondear un número hacia arriba hasta el próximo número entero. (La Función FLOOR redondea un número hacia abajo hasta el próximo número entero).

### Sintaxis

```
{CEIL | CEILING}(number)
```

### Argumentos

número

El número o la expresión que toma el valor de un número. Puede ser el tipo SMALLINT, INTEGER, BIGINT, DECIMAL, FLOAT4, FLOAT8 o SUPER.

### Tipo de retorno

CEILING y CEIL devuelven el mismo tipo de datos como su argumento.

Cuando la entrada es del tipo SUPER, la salida conserva el mismo tipo dinámico que la entrada mientras que el tipo estático sigue siendo el tipo SUPER. Cuando el tipo dinámico de SUPER no es un número, Amazon Redshift devuelve un valor nulo.

### Ejemplos

El siguiente ejemplo utiliza la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para calcular el límite máximo de la comisión pagada para una transacción de ventas dada, utilice el siguiente ejemplo.

```
SELECT CEILING(commission) FROM sales
WHERE salesid=10000;

+---------+
| ceiling |
+---------+
|      29 |
```

```
+---------+
```

## Función COS

COS es una función trigonométrica que devuelve el coseno de un número. El valor de retorno está en radianes y se encuentra entre `-1` y `1`, inclusive.

### Sintaxis

```
COS(double_precision)
```

### Argumentos

número

El parámetro de entrada es un número de `DOUBLE PRECISION`.

### Tipo de retorno

La función COS devuelve un número de `DOUBLE PRECISION`.

### Ejemplos

Para devolver el coseno de `0`, use el siguiente ejemplo.

```
SELECT COS(0);

+-----+
| cos |
+-----+
|   1 |
+-----+
```

Para devolver el coseno de `pi`, use el siguiente ejemplo.

```
SELECT COS(PI());

+-----+
| cos |
+-----+
|  -1 |
```

```
+-----+
```

## Función COT

COT es una función trigonométrica que devuelve la cotangente de un número. El parámetro de entrada debe ser distinto de cero.

### Sintaxis

```
COT(number)
```

### Argumento

número

El parámetro de entrada es un número de DOUBLE PRECISION.

### Tipo de retorno

DOUBLE PRECISION

### Ejemplos

Para devolver la cotangente de 1, use el siguiente ejemplo.

```
SELECT COT(1);

+--------------------+
|        cot         |
+--------------------+
| 0.6420926159343306 |
+--------------------+
```

## Función DEGREES

Convierte un ángulo en radianes a su equivalente en grados.

### Sintaxis

```
DEGREES(number)
```

Argumento

número

El parámetro de entrada es un número de DOUBLE PRECISION.

Tipo de retorno

DOUBLE PRECISION

Ejemplos

Para devolver el equivalente en grados de .5 radianes, use el siguiente ejemplo.

```
SELECT DEGREES(.5);

+-------------------+
|      degrees      |
+-------------------+
| 28.64788975654116 |
+-------------------+
```

Para convertir radianes de Pi a grados, use el siguiente ejemplo.

```
SELECT DEGREES(pi());

+---------+
| degrees |
+---------+
|     180 |
+---------+
```

## Función DEXP

La función DEXP devuelve el valor exponential en notación científica para un número de doble precisión. La única diferencia entre las funciones DEXP y EXP es que el parámetro para DEXP debe ser de DOUBLE PRECISION.

Sintaxis

```
DEXP(number)
```

Argumento

número

El parámetro de entrada es un número de `DOUBLE PRECISION`.

Tipo de retorno

`DOUBLE PRECISION`

Ejemplo

El siguiente ejemplo utiliza la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Se utiliza la función DEXP para prever las ventas de tickets según un patrón de crecimiento continuo. En este ejemplo, la subconsulta devuelve la cantidad de tickets vendidos en 2008. El resultado se multiplica por el resultado de la función DEXP, que especifica un índice de crecimiento continuo del 7% durante 10 años.

```
SELECT (SELECT SUM(qtysold)
FROM sales, date
WHERE sales.dateid=date.dateid
AND year=2008) * DEXP((7::FLOAT/100)*10) qty2010;

+-------------------+
|      qty2010      |
+-------------------+
| 695447.4837722216 |
+-------------------+
```

## Función DLOG1

La función DLOG1 devuelve el logaritmo natural del parámetro de entrada. Sinónimo de Función LN.

## Función DLOG10

DLOG10 devuelve el logaritmo de base 10 del parámetro de entrada.

Sinónimo de Función LOG.

Sintaxis

```
DLOG10(number)
```

Argumento

número

El parámetro de entrada es un número de DOUBLE PRECISION.

Tipo de retorno

DOUBLE PRECISION

Ejemplo

Para devolver el logaritmo de base 10 del número 100, use el siguiente ejemplo.

```
SELECT DLOG10(100);

+--------+
| dlog10 |
+--------+
|      2 |
+--------+
```

## Función EXP

La función EXP implementa la función exponencial para una expresión numérica, o la base del logaritmo natural, e, elevada a potencia de expresión. La función EXP es la operación inversa de Función LN.

Sintaxis

```
EXP(expression)
```

Argumento

expresión

La expresión de entrada debe ser un tipo de datos INTEGER, DECIMAL o DOUBLE PRECISION.

Tipo de retorno

`DOUBLE PRECISION`

Ejemplo

El siguiente ejemplo utiliza la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Se utiliza la función EXP para prever las ventas de tickets según un patrón de crecimiento continuo. En este ejemplo, la subconsulta devuelve la cantidad de tickets vendidos en 2008. El resultado se multiplica por el resultado de la función EXP, que especifica un índice de crecimiento continuo del 7% durante 10 años.

```
SELECT (SELECT SUM(qtysold)
FROM sales, date
WHERE sales.dateid=date.dateid
AND year=2008) * EXP((7::FLOAT/100)*10) qty2018;

+-------------------+
|      qty2018      |
+-------------------+
| 695447.4837722216 |
+-------------------+
```

## Función FLOOR

La función FLOOR redondea un número hacia abajo hasta el próximo número entero.

Sintaxis

```
FLOOR(number)
```

Argumento

número

El número o la expresión que toma el valor de un número. Puede ser el tipo `SMALLINT`, `INTEGER`, `BIGINT`, `DECIMAL`, `FLOAT4`, `FLOAT8` o `SUPER`.

Tipo de retorno

FLOOR devuelve el mismo tipo de datos como su argumento.

Cuando la entrada es del tipo SUPER, la salida conserva el mismo tipo dinámico que la entrada mientras que el tipo estático sigue siendo el tipo SUPER. Cuando el tipo dinámico de SUPER no es un número, Amazon Redshift devuelve NULL.

Ejemplos

En los siguientes ejemplos se utiliza la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para mostrar el valor de la comisión pagada por una transacción de ventas determinada antes y después de usar la función FLOOR, use el siguiente ejemplo.

```
SELECT commission
FROM sales
WHERE salesid=10000;

+------------+
| commission |
+------------+
|      28.05 |
+------------+

SELECT FLOOR(commission)
FROM sales
WHERE salesid=10000;

+-------+
| floor |
+-------+
|    28 |
+-------+
```

## Función LN

Devuelve el logaritmo natural del parámetro de entrada.

Sinónimo de Función DLOG1.

### Sintaxis

```
LN(expression)
```

### Argumento

expresión

La columna o expresión de destino sobre la que opera la función.

> **Note**
>
> Esta función devuelve un error para algunos tipos de datos si la expresión referencia una tabla de Amazon Redshift creada por el usuario o a una tabla de sistema STL o STV de Amazon Redshift.

Las expresiones con los siguientes tipos de datos producen un error si usa como referencia una tabla de sistema o creada por usuarios. Las expresiones con estos tipos de datos se ejecutan exclusivamente en el nodo principal:

- `BOOLEAN`
- `CHAR`
- `DATE`
- `DECIMAL` o `NUMERIC`
- `TIMESTAMP`
- `VARCHAR`

Las expresiones con los siguientes tipos de datos se ejecutan con éxito en tablas creadas por usuarios y tablas de sistema STL o STV:

- `BIGINT`
- `DOUBLE PRECISION`
- `INTEGER`
- `REAL`
- `SMALLINT`

Tipo de retorno

La función LN devuelve el mismo tipo que la expresión de entrada.

Ejemplos

Para devolver el logaritmo natural o la base de logaritmo e del número 2.718281828, use el siguiente ejemplo.

```
SELECT LN(2.718281828);

+--------------------+
|         ln         |
+--------------------+
| 0.9999999998311267 |
+--------------------+
```

Tenga en cuenta que la respuesta es casi igual a 1.

El siguiente ejemplo utiliza la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para devolver el logaritmo natural de los valores en la columna userid en la tabla USERS, use el ejemplo siguiente.

```
SELECT username, LN(userid) FROM users ORDER BY userid LIMIT 10;

+----------+--------------------+
| username |         ln         |
+----------+--------------------+
| JSG99FHE |                  0 |
| PGL08LJI | 0.6931471805599453 |
| IFT66TXU | 1.0986122886681098 |
| XDZ38RDD | 1.3862943611198906 |
| AEB55QTM | 1.6094379124341003 |
| NDQ15VBM |  1.791759469228055 |
| OWY35QYB | 1.9459101490553132 |
| AZG78YIP | 2.0794415416798357 |
| MSD36KVR | 2.1972245773362196 |
| WKW41AIW |  2.302585092994046 |
+----------+--------------------+
```

## Función LOG

Devuelve el logaritmo de un número.

Si utiliza esta función para calcular el logaritmo en base 10, también puede usar Función DLOG10.

### Sintaxis

```
LOG([base, ]argument)
```

### Parámetros

base

(Opcional) La base de la función del logaritmo. Este número debe ser positivo y no puede ser igual a 1. Si se omite este parámetro, Amazon Redshift calculará el logaritmo en base 10 del argumento.

argument

El argumento de la función logarítmica. Este número debe ser positivo. Si el valor del argumento es 1, la función devuelve `0`.

### Tipo de retorno

La función LOG devuelve un número de `DOUBLE PRECISION`.

### Ejemplos

Para encontrar el logaritmo en base 2 de 100, utilice el siguiente ejemplo.

```
SELECT LOG(2, 100);
+-------------------+
|        log        |
+-------------------+
| 6.643856189774725 |
+-------------------+
```

Para encontrar el logaritmo en base 10 de 100, utilice el siguiente ejemplo. Tenga en cuenta que si omite el parámetro base, Amazon Redshift asume una base de 10.

```
SELECT LOG(100);
```

```
+-----+
| log |
+-----+
|   2 |
+-----+
```

## Función MOD

Devuelve el resto de dos números, también denominada operación de módulo. Para calcular el resultado, el primer parámetro se divide entre el segundo.

### Sintaxis

```
MOD(number1, number2)
```

### Argumentos

number1

El primer parámetro de entrada es un número `INTEGER`, `SMALLINT`, `BIGINT` o `DECIMAL`. Si cada parámetro es de tipo `DECIMAL`, el otro parámetro debe ser también un tipo `DECIMAL`. Si cada parámetro es un valor `INTEGER`, el otro parámetro puede ser un valor `INTEGER`, `SMALLINT` o `BIGINT`. Los parámetros pueden también ser `SMALLINT` o `BIGINT`, pero un parámetro no puede ser `SMALLINT` si el otro es `BIGINT`.

number2

El segundo parámetro es un número `INTEGER`, `SMALLINT`, `BIGINT` o `DECIMAL`. Se aplican las mismas reglas de tipo de datos en number2 y en number1.

### Tipo de retorno

El tipo de retorno de la función MOD es el mismo tipo numérico que los parámetros de entrada, si ambos parámetros de entrada son del mismo tipo. No obstante, si algún parámetro de entrada es un valor `INTEGER`, el tipo de retorno también será `INTEGER`. Los tipos de devolución válidos son `DECIMAL`, `INT`, `SMALLINT` y `BIGINT`.

### Notas de uso

Puede utilizar `%` como operador de módulo.

## Ejemplos

Para devolver el resto cuando se divide un número entre otro, use el siguiente ejemplo.

```
SELECT MOD(10, 4);

+-----+
| mod |
+-----+
|   2 |
+-----+
```

Para devolver un resultado `DECIMAL` al utilizar la función MOD, utilice el siguiente ejemplo.

```
SELECT MOD(10.5, 4);

+-----+
| mod |
+-----+
| 2.5 |
+-----+
```

Para convertir un número antes de ejecutar la función MOD, utilice el siguiente ejemplo. Para obtener más información, consulte Función CAST.

```
SELECT MOD(CAST(16.4 AS INTEGER), 5);

+-----+
| mod |
+-----+
|   1 |
+-----+
```

Para comprobar si el primer parámetro es par dividiéndolo entre 2, use el siguiente ejemplo.

```
SELECT mod(5,2) = 0 AS is_even;

+---------+
| is_even |
+---------+
| false   |
```

```
+---------+
```

Para usar % como operador de módulo, utilice el siguiente ejemplo.

```
SELECT 11 % 4 as remainder;

 +-----------+
| remainder |
+-----------+
|         3 |
+-----------+
```

El siguiente ejemplo utiliza la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para devolver la información para categorías con números impares en la tabla CATEGORY, use el siguiente ejemplo.

```
SELECT catid, catname
FROM category
WHERE MOD(catid,2)=1
ORDER BY 1,2;

+-------+-----------+
| catid |  catname  |
+-------+-----------+
|     1 | MLB       |
|     3 | NFL       |
|     5 | MLS       |
|     7 | Plays     |
|     9 | Pop       |
|    11 | Classical |
+-------+-----------+
```

## Función PI

La función PI devuelve el valor de Pi a 14 lugares decimales.

### Sintaxis

```
PI()
```

Tipo de retorno

DOUBLE PRECISION

Ejemplos

Para devolver el valor de pi, utilice el ejemplo siguiente.

```
SELECT PI();

+-------------------+
|        pi         |
+-------------------+
| 3.141592653589793 |
+-------------------+
```

## Función POWER

La función POWER es una función exponencial que eleva una expresión numérica a la potencia de una segunda expresión numérica. Por ejemplo, 2 a la tercera potencia se calcula como POWER(2,3), con un resultado de 8.

Sintaxis

```
{POW | POWER}(expression1, expression2)
```

Argumentos

expression1

Expresión numérica que se elevará. Debe ser un tipo de datos INTEGER, DECIMAL o FLOAT.

expression2

Potencia a la que se va a elevar expression1. Debe ser un tipo de datos INTEGER, DECIMAL o FLOAT.

Tipo de retorno

DOUBLE PRECISION

### Ejemplos

En los siguientes ejemplos se utiliza la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

En el siguiente ejemplo, la función POWER se usa para prever cómo serán las ventas de tickets en los próximos 10 años, según la cantidad de tickets vendidos en 2008 (el resultado de la subconsulta). El índice de crecimiento se establece en 7% por año en este ejemplo.

```
SELECT (SELECT SUM(qtysold) FROM sales, date
WHERE sales.dateid=date.dateid
AND year=2008) * POW((1+7::FLOAT/100),10) qty2010;

+-------------------+
|      qty2010      |
+-------------------+
| 679353.7540885945 |
+-------------------+
```

El siguiente ejemplo es una variación del ejemplo anterior, con el índice de crecimiento en 7 % por año, pero con el intervalo establecido en meses (120 meses en 10 años).

```
SELECT (SELECT SUM(qtysold) FROM sales, date
WHERE sales.dateid=date.dateid
AND year=2008) * POW((1+7::FLOAT/100/12),120) qty2010;

+-----------------+
|     qty2010     |
+-----------------+
| 694034.54678046 |
+-----------------+
```

## Función RADIANS

La función RADIANS convierte un ángulo en grados a su equivalente en radianes.

### Sintaxis

```
RADIANS(number)
```

Argumento

número

El parámetro de entrada es un número de DOUBLE PRECISION.

Tipo de retorno

DOUBLE PRECISION

Ejemplos

Para devolver el equivalente en radianes de 180 grados, use el siguiente ejemplo.

```
SELECT RADIANS(180);

+-------------------+
|      radians      |
+-------------------+
| 3.141592653589793 |
+-------------------+
```

## Función RANDOM

La función RANDOM genera un valor aleatorio entre 0,0 (inclusive) y 1,0 (exclusive).

Sintaxis

```
RANDOM()
```

Tipo de retorno

DOUBLE PRECISION

Notas de uso

Ejecute RANDOM luego de establecer un valor de inicio con el comando SET para que RANDOM genere números, en una secuencia predecible.

Ejemplos

Para calcular un valor aleatorio entre 0 y 99, utilice el siguiente ejemplo. Si el número aleatorio está comprendido entre 0 y 1, esta consulta produce un número aleatorio comprendido entre 0 y 100.

```
SELECT CAST(RANDOM() * 100 AS INT);

+------+
| int4 |
+------+
|   59 |
+------+
```

Este ejemplo usa el comando SET para establecer un valor SEED de modo que RANDOM genere una secuencia predecible de números.

Para devolver tres enteros RANDOM sin establecer el valor SEED, utilice el siguiente ejemplo.

```
SELECT CAST(RANDOM() * 100 AS INT);
+------+
| int4 |
+------+
|    6 |
+------+

SELECT CAST(RANDOM() * 100 AS INT);
+------+
| int4 |
+------+
|   68 |
+------+

SELECT CAST(RANDOM() * 100 AS INT);
+------+
| int4 |
+------+
|   56 |
+------+
```

Para establecer el valor SEED en `.25` y devolver tres números RANDOM más, utilice el siguiente ejemplo.

```
SET SEED TO .25;
SELECT CAST(RANDOM() * 100 AS INT);
+------+
| int4 |
+------+
```

```
|   21 |
+------+

SELECT CAST(RANDOM() * 100 AS INT);
+------+
| int4 |
+------+
|   79 |
+------+

SELECT CAST(RANDOM() * 100 AS INT);
+------+
| int4 |
+------+
|   12 |
+------+
```

Para restablecer el valor SEED a `.25` y verificar que RANDOM devuelva los mismos resultados que en las tres llamadas anteriores, use el siguiente ejemplo.

```
SET SEED TO .25;
SELECT CAST(RANDOM() * 100 AS INT);
+------+
| int4 |
+------+
|   21 |
+------+

SELECT CAST(RANDOM() * 100 AS INT);
+------+
| int4 |
+------+
|   79 |
+------+

SELECT CAST(RANDOM() * 100 AS INT);
+------+
| int4 |
+------+
|   12 |
+------+
```

En los siguientes ejemplos se utiliza la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para recuperar un ejemplo aleatorio uniforme de 10 artículos de la tabla SALES, utilice el siguiente ejemplo.

```
SELECT *
FROM sales
ORDER BY RANDOM()
LIMIT 10;

+---------+--------+----------+---------+---------+--------+---------+-----------
+------------+---------------------+
| salesid | listid | sellerid | buyerid | eventid | dateid | qtysold | pricepaid |
 commission |      saletime       |
+---------+--------+----------+---------+---------+--------+---------+-----------
+------------+---------------------+
|   45422 |  51114 |     5983 |   24482 |    4369 |   2118 |       1 |       195 |
 29.25 | 2008-10-19 05:20:07 |
|   42481 |  47638 |     4573 |    6198 |    6479 |   1987 |       4 |      1140 |
   171 | 2008-06-10 09:39:19 |
|   31494 |  34759 |    18895 |    4719 |    7753 |   2090 |       4 |      1024 |
 153.6 | 2008-09-21 03:44:26 |
|  119388 | 136685 |    21815 |   41905 |    2071 |   1884 |       1 |       359 |
 53.85 | 2008-02-27 10:43:10 |
|  166990 | 225037 |    18529 |    7628 |     746 |   2113 |       1 |      2009 |
 301.35 | 2008-10-14 10:07:44 |
|   11146 |  12096 |    42685 |    6619 |    1876 |   2123 |       1 |        29 |
  4.35 | 2008-10-24 06:23:54 |
|  148537 | 172056 |    15102 |   11787 |    6122 |   1923 |       2 |       480 |
    72 | 2008-04-07 03:58:23 |
|   68945 |  78387 |     7359 |   18323 |    6636 |   1910 |       1 |       457 |
 68.55 | 2008-03-25 08:31:03 |
|   52796 |  59576 |     9909 |   15102 |    7958 |   1951 |       1 |       479 |
 71.85 | 2008-05-05 02:25:08 |
|   90684 | 103522 |    38052 |   21549 |    7384 |   2117 |       1 |       313 |
 46.95 | 2008-10-18 05:43:11 |
+---------+--------+----------+---------+---------+--------+---------+-----------
+------------+---------------------+
```

Para recuperar un ejemplo aleatorio de 10 objetos, pero elegir los objetos en proporción a sus precios, use el siguiente ejemplo. Por ejemplo, un objeto que cuesta el doble del precio de otro tendría el doble de posibilidades de aparecer en los resultados de la búsqueda.

```
SELECT *
FROM sales
ORDER BY -LOG(RANDOM()) / pricepaid
LIMIT 10;

+---------+--------+----------+---------+---------+--------+---------+-----------
+------------+---------------------+
| salesid | listid | sellerid | buyerid | eventid | dateid | qtysold | pricepaid |
 commission |      saletime       |
+---------+--------+----------+---------+---------+--------+---------+-----------
+------------+---------------------+
|  158340 | 208208 |    17082 |   42018 |    1211 |   2160 |       4 |      6852 |
 1027.8 | 2008-11-30 12:21:43 |
|   53250 |  60069 |    12644 |    7066 |    7942 |   1838 |       4 |      1528 |
 229.2 | 2008-01-12 11:24:56 |
|   22929 |  24938 |    47314 |    6503 |     179 |   2000 |       3 |       741 |
 111.15 | 2008-06-23 08:04:50 |
|  164980 | 221181 |     1949 |   19670 |    1471 |   1906 |       1 |      1330 |
 199.5 | 2008-03-21 07:59:51 |
|  159641 | 211179 |    44897 |   16652 |    7458 |   2128 |       1 |      1019 |
 152.85 | 2008-10-29 02:02:15 |
|   73143 |  83439 |     5716 |    5727 |    7314 |   1903 |       1 |       248 |
  37.2 | 2008-03-18 11:07:42 |
|   84778 |  96749 |    46608 |   32980 |    3883 |   1999 |       2 |       958 |
 143.7 | 2008-06-22 12:13:31 |
|  171096 | 232929 |    43683 |    8536 |    8353 |   1870 |       1 |       929 |
 139.35 | 2008-02-13 01:36:36 |
|   74212 |  84697 |    39809 |   15569 |    5525 |   2105 |       2 |       896 |
 134.4 | 2008-10-06 11:47:50 |
|  158011 | 207556 |    25399 |   16881 |     232 |   2088 |       2 |      2526 |
 378.9 | 2008-09-19 06:00:26 |
+---------+--------+----------+---------+---------+--------+---------+-----------
+------------+---------------------+
```

## Función ROUND

La función ROUND redondea los números hasta el valor entero o decimal más cercano.

La función ROUND puede incluir, de forma opcional, un segundo argumento como un valor `INTEGER` que indique la cantidad de lugares decimales para el redondeo, sea cual sea la dirección. Cuando no se proporciona el segundo argumento, la función redondea al número entero más cercano. Cuando

se especifica el segundo argumento integer, la función redondea al número más cercano con una precisión de hasta integer decimales.

Sintaxis

```
ROUND(number [ , integer ] )
```

Argumentos

número

Un número o una expresión que toma el valor de un número. Puede ser el tipo DECIMAL, FLOAT8 o SUPER. Amazon Redshift puede convertir implícitamente otros tipos de datos numéricos.

integer

(Opcional) Un número INTEGER que indica la cantidad de lugares decimales para el redondeo, sea cual sea la dirección. El tipo de datos SUPER no se admite para este argumento.

Tipo de retorno

ROUND devuelve el mismo tipo de datos numéricos como el número de entrada.

Cuando la entrada es del tipo SUPER, la salida conserva el mismo tipo dinámico que la entrada mientras que el tipo estático sigue siendo el tipo SUPER. Cuando el tipo dinámico de SUPER no es un número, Amazon Redshift devuelve NULL.

Ejemplos

En los siguientes ejemplos se utiliza la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para redondear la comisión pagada para una transacción dada hasta el número entero más cercano, use el siguiente ejemplo.

```
SELECT commission, ROUND(commission)
FROM sales WHERE salesid=10000;

+------------+-------+
| commission | round |
+------------+-------+
```

```
|      28.05 |    28 |
+------------+-------+
```

Para redondear la comisión pagada para una transacción dada hasta el primer lugar decimal, use el siguiente ejemplo.

```
SELECT commission, ROUND(commission, 1)
FROM sales WHERE salesid=10000;

+------------+-------+
| commission | round |
+------------+-------+
|      28.05 |  28.1 |
+------------+-------+
```

Para ampliar la precisión en la dirección opuesta a la del ejemplo anterior, utilice el ejemplo siguiente.

```
SELECT commission, ROUND(commission, -1)
FROM sales WHERE salesid=10000;

+------------+-------+
| commission | round |
+------------+-------+
|      28.05 |    30 |
+------------+-------+
```

## Función SIN

SIN es una función trigonométrica que devuelve el seno de un número. El valor devuelto está comprendido entre -1 y 1.

### Sintaxis

```
SIN(number)
```

### Argumento

número

Un número de DOUBLE PRECISION en radianes.

Tipo de retorno

`DOUBLE PRECISION`

Ejemplos

Para devolver el seno de `-PI`, use el siguiente ejemplo.

```
SELECT SIN(-PI());

+-------------------------+
|           sin           |
+-------------------------+
| -0.00000000000000012246 |
+-------------------------+
```

## Función SIGN

La función SIGN devuelve el signo (positivo o negativo) de un número. El resultado de la función SIGN es 1 si el argumento es positivo, `-1` si el argumento es negativo o `0` si el argumento es `0`.

Sintaxis

```
SIGN(number)
```

Argumento

número

Número o expresión que toma el valor de un número. Puede ser un tipo `DECIMAL`, `FLOAT8` o `SUPER`. Amazon Redshift puede convertir otros tipos de datos de acuerdo con las reglas de conversión implícitas.

Tipo de retorno

SIGN devuelve el mismo tipo de datos numérico como el argumento de entrada. Si la entrada es `DECIMAL`, el resultado es `DECIMAL(1,0)`.

Cuando la entrada es del tipo `SUPER`, la salida conserva el mismo tipo dinámico que la entrada mientras que el tipo estático sigue siendo el tipo `SUPER`. Cuando el tipo dinámico de `SUPER` no es un número, Amazon Redshift devuelve `NULL`.

### Ejemplos

En el siguiente ejemplo, se muestra que la columna d en la tabla t2 tiene `DOUBLE PRECISION` como su tipo, ya que la entrada es `DOUBLE PRECISION` y esa columna n en la tabla t2 tiene `NUMERIC(1,0)` como la salida ya que la entrada es `NUMERIC`.

```
CREATE TABLE t1(d DOUBLE PRECISION, n NUMERIC(12, 2));
INSERT INTO t1 VALUES (4.25, 4.25), (-4.25, -4.25);
CREATE TABLE t2 AS SELECT SIGN(d) AS d, SIGN(n) AS n FROM t1;
SELECT table_name, column_name, data_type FROM SVV_REDSHIFT_COLUMNS WHERE
 table_name='t1' OR table_name='t2';

+------------+-------------+----------------------+
| table_name | column_name |       data_type      |
+------------+-------------+----------------------+
| t1         | d           | double precision     |
| t1         | n           | numeric(12,2)        |
| t2         | d           | double precision     |
| t2         | n           | numeric(1,0)         |
| t1         | col1        | character varying(20) |
+------------+-------------+----------------------+
```

El siguiente ejemplo utiliza la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para determinar el signo de la comisión pagada por una transacción determinada a partir de la tabla SALES, utilice el siguiente ejemplo.

```
SELECT commission, SIGN(commission)
FROM sales WHERE salesid=10000;

+------------+------+
| commission | sign |
+------------+------+
|      28.05 |    1 |
+------------+------+
```

## Función SQRT

La función SQRT devuelve la raíz cuadrada de un valor `NUMERIC`. La raíz cuadrada es un número multiplicado por sí mismo para obtener el valor dado.

### Sintaxis

```
SQRT(expression)
```

### Argumento

expresión

La expresión debe tener un tipo de datos `INTEGER`, `DECIMAL` o `FLOAT` o un tipo de datos que se convierte implícitamente en esos tipos de datos. La expresión puede incluir funciones.

### Tipo de retorno

`DOUBLE PRECISION`

### Ejemplos

Para devolver la raíz cuadrada de 16, use el siguiente ejemplo.

```
SELECT SQRT(16);

+------+
| sqrt |
+------+
|    4 |
+------+
```

Para devolver la raíz cuadrada de la cadena 16 mediante una conversión de tipos implícita, utilice el siguiente ejemplo.

```
SELECT SQRT('16');

+------+
| sqrt |
+------+
|    4 |
+------+
```

Para devolver la raíz cuadrada de 16,4 después de usar la función ROUND, utilice el siguiente ejemplo.

```
SELECT SQRT(ROUND(16.4));

+------+
| sqrt |
+------+
|    4 |
+------+
```

Para devolver la longitud del radio si se da el área de un círculo, use el siguiente ejemplo. Calcula el radio en pulgadas, por ejemplo, cuando se le da el área en pulgadas cuadradas. El área del ejemplo es 20.

```
SELECT SQRT(20/PI()) AS radius;

+--------------------+
|      radius        |
+--------------------+
| 2.5231325220201604 |
+--------------------+
```

En los siguientes ejemplos se utiliza la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para devolver la raíz cuadrada para valores COMMISSION de la tabla SALES, use el siguiente ejemplo. La columna COMMISSION es una columna `DECIMAL`. En este ejemplo se muestra cómo se puede utilizar la función en una consulta con una lógica condicional más compleja.

```
SELECT SQRT(commission)
FROM sales WHERE salesid < 10 ORDER BY salesid;

+--------------------+
|        sqrt        |
+--------------------+
| 10.449880382090505 |
| 3.3763886032268267 |
|  7.245688373094719 |
|  5.123475382979799 |
|  4.806245936279167 |
|  7.687652437513028 |
| 10.871982339941507 |
| 5.4359911699707535 |
```

```
|   9.41541289588513 |
+--------------------+
```

Para devolver la raíz cuadrada redondeada para el mismo conjunto de valores COMMISSION, use el siguiente ejemplo.

```
SELECT ROUND(SQRT(commission))
FROM sales WHERE salesid < 10 ORDER BY salesid;

+-------+
| round |
+-------+
|    10 |
|     3 |
|     7 |
|     5 |
|     5 |
|     8 |
|    11 |
|     5 |
|     9 |
+-------+
```

## Función TAN

TAN es una función trigonométrica que devuelve la tangente de un número. El argumento de entrada es un número (en radianes).

Sintaxis

```
TAN(number)
```

Argumento

número

Un número de DOUBLE PRECISION.

Tipo de retorno

DOUBLE PRECISION

### Ejemplos

Para devolver la tangente de cero, utilice el siguiente ejemplo.

```
SELECT TAN(0);

+-----+
| tan |
+-----+
|   0 |
+-----+
```

## Función TRUNC

La función TRUNC trunca los números hasta el valor entero o decimal anterior.

La función TRUNC puede incluir, de forma opcional, un segundo argumento como un valor `INTEGER` que indique la cantidad de lugares decimales para el redondeo, sea cual sea la dirección. Cuando no se proporciona el segundo argumento, la función redondea al número entero más cercano. Cuando se especifica el segundo argumento integer, la función redondea al número más cercano con una precisión de hasta integer decimales.

Esta función también puede truncar un `TIMESTAMP` y devolver una `DATE`. Para obtener más información, consulte Función TRUNC.

### Sintaxis

```
TRUNC(number [ , integer ])
```

### Argumentos

número

Un número o una expresión que toma el valor de un número. Puede ser el tipo `DECIMAL`, `FLOAT8` o `SUPER`. Amazon Redshift puede convertir otros tipos de datos de acuerdo con las reglas de conversión implícitas.

integer

(Opcional) Un número `INTEGER` que indica la cantidad de lugares decimales de precisión, sea cual sea la dirección. Si no se proporciona un valor entero, el número se trunca como un

número entero; si se especifica un número entero, el número se trunca hasta el lugar decimal especificado. Esto no es compatible con el tipo de datos SUPER.

Tipo de retorno

TRUNC devuelve el mismo tipo de datos que el número de entrada.

Cuando la entrada es del tipo SUPER, la salida conserva el mismo tipo dinámico que la entrada mientras que el tipo estático sigue siendo el tipo SUPER. Cuando el tipo dinámico de SUPER no es un número, Amazon Redshift devuelve NULL.

Ejemplos

Algunos de los ejemplos siguientes utilizan la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para truncar la comisión pagada para una transacción dada de ventas, utilice el siguiente ejemplo.

```
SELECT commission, TRUNC(commission)
FROM sales WHERE salesid=784;

+------------+-------+
| commission | trunc |
+------------+-------+
|     111.15 |   111 |
+------------+-------+
```

Para truncar el mismo valor de comisión hasta el primer lugar decimal, utilice el siguiente ejemplo.

```
SELECT commission, TRUNC(commission,1)
FROM sales WHERE salesid=784;

+------------+-------+
| commission | trunc |
+------------+-------+
|     111.15 | 111.1 |
+------------+-------+
```

Para truncar la comisión con un valor negativo para el segundo argumento, utilice el siguiente ejemplo. Tenga en cuenta que 111.15 se redondea a 110.

```
SELECT commission, TRUNC(commission,-1)
FROM sales WHERE salesid=784;

+------------+-------+
| commission | trunc |
+------------+-------+
|     111.15 |   110 |
+------------+-------+
```

## Funciones de objeto

A continuación, se presentan las funciones de objeto SQL que Amazon Redshift admite para crear objetos de tipo SUPER:

Temas

- Función LOWER_ATTRIBUTE_NAMES
- Función OBJECT
- Función OBJECT_TRANSFORM
- Función UPPER_ATTRIBUTE_NAMES

### Función LOWER_ATTRIBUTE_NAMES

Convierte todos los nombres de atributos aplicables de un valor SUPER a minúsculas, con la misma rutina de conversión de mayúsculas y minúsculas que Función LOWER. LOWER_ATTRIBUTE_NAMES admite caracteres multibyte UTF-8 de hasta un máximo de cuatro bytes por carácter.

Para convertir los nombres del atributo SUPER a mayúsculas, use Función UPPER_ATTRIBUTE_NAMES.

Sintaxis

```
LOWER_ATTRIBUTE_NAMES(super_expression)
```

Argumentos

super_expression

Una expresión SUPER.

Tipo de retorno

SUPER

Notas de uso

En Amazon Redshift, los identificadores de columna tradicionalmente no distinguen entre mayúsculas y minúsculas y se convierten a minúsculas. Si ingiere datos de formatos de datos que distinguen mayúsculas de minúsculas, como JSON, es posible que los datos contengan nombres de atributos con mayúsculas y minúsculas.

Considere el siguiente ejemplo.

```
CREATE TABLE t1 (s) AS SELECT JSON_PARSE('{"AttributeName": "Value"}');


SELECT s.AttributeName FROM t1;

attributename
-------------
NULL


SELECT s."AttributeName" FROM t1;

attributename
-------------
NULL
```

Amazon Redshift devuelve un valor NULL para ambas consultas. Para realizar consultas de `AttributeName`, utilice LOWER_ATTRIBUTE_NAMES para convertir los nombres de los atributos de los datos a minúsculas. Considere el siguiente ejemplo.

```
CREATE TABLE t2 (s) AS SELECT LOWER_ATTRIBUTE_NAMES(s) FROM t1;


SELECT s.attributename FROM t2;

attributename
-------------
"Value"
```

```
SELECT s.AttributeName FROM t2;

attributename
-------------
"Value"


SELECT s."attributename" FROM t2;

attributename
-------------
"Value"


SELECT s."AttributeName" FROM t2;

attributename
-------------
"Value"
```

Una opción relacionada para trabajar con nombres de atributos de objetos en mayúsculas y minúsculas es la opción de configuración `enable_case_sensitive_super_attribute`, que permite a Amazon Redshift reconocer mayúsculas y minúsculas en los nombres de atributo SUPER. Esta puede ser una solución alternativa al uso de LOWER_ATTRIBUTE_NAMES. Para obtener más información sobre `enable_case_sensitive_super_attribute`, diríjase a enable_case_sensitive_super_attribute.

Ejemplos

Conversión de los nombres de los atributos SUPER a minúsculas

En el siguiente ejemplo, se utiliza LOWER_ATTRIBUTE_NAMES para convertir los nombres de los atributos de todos los valores SUPER de una tabla.

```
-- Create a table and insert several SUPER values.
CREATE TABLE t (i INT, s SUPER);

INSERT INTO t VALUES
  (1, NULL),
  (2, 'A'::SUPER),
  (3, JSON_PARSE('{"AttributeName": "B"}')),
```

```
  (4, JSON_PARSE(
     '[{"Subobject": {"C": "C"},
       "Subarray": [{"D": "D"}, "E"]
      }]'));

-- Convert all attribute names to lowercase.
UPDATE t SET s = LOWER_ATTRIBUTE_NAMES(s);

SELECT i, s FROM t ORDER BY i;

 i |                        s
---+-------------------------------------------------
 1 | NULL
 2 | "A"
 3 | {"attributename":"B"}
 4 | [{"subobject":{"c":"C"},"subarray":[{"d":"D"}, "E"]}]
```

Observe cómo funciona LOWER_ATTRIBUTE_NAMES.

- Los valores NULL y los valores escalares SUPER como `"A"` permanecen inalterados.
- En un objeto SUPER, todos los nombres de los atributos se cambian a minúsculas, pero los valores de los atributos, como `"B"`, permanecen inalterados.
- LOWER_ATTRIBUTE_NAMES se aplica de forma recursiva a cualquier objeto SUPER que esté anidado dentro de una matriz SUPER o dentro de otro objeto.

Uso de LOWER_ATTRIBUTE_NAMES en un objeto SUPER con nombres de atributos duplicados

Si un objeto SUPER contiene atributos cuyos nombres solo difieren en las mayúsculas y minúsculas, LOWER_ATTRIBUTE_NAMES generará un error. Considere el siguiente ejemplo.

```
SELECT LOWER_ATTRIBUTE_NAMES(JSON_PARSE('{"A": "A", "a": "a"}'));

error:   Invalid input
code:    8001
context: SUPER value has duplicate attributes after case conversion.
```

## Función OBJECT

Crea un objeto del tipo de datos SUPER.

### Sintaxis

```
OBJECT ( [ key1, value1 ], [ key2, value2 ...] )
```

### Argumentos

key1, key2

Expresiones que se evalúan como cadenas de tipo VARCHAR.

value1, value2

Expresiones de cualquier tipo de datos de Amazon Redshift, excepto los datetime, ya que Amazon Redshift no convierte los tipos de datetime al tipo de datos SUPER. Para obtener más información sobre los tipos de datetime, consulte Tipos de fecha y hora.

Las expresiones `value` en un objeto no tienen que ser del mismo tipo de datos.

### Tipo de retorno

SUPER

### Ejemplo

```
-- Creates an empty object.
select object();

object
--------
{}
(1 row)

-- Creates objects with different keys and values.
select object('a', 1, 'b', true, 'c', 3.14);

object
---------------------------
{"a":1,"b":true,"c":3.14}
(1 row)

select object('a', object('aa', 1), 'b', array(2,3), 'c', json_parse('{}'));
```

```
object
---------------------------------
{"a":{"aa":1},"b":[2,3],"c":{}}
(1 row)

-- Creates objects using columns from a table.
create table bar (k varchar, v super);
insert into bar values ('k1', json_parse('[1]')), ('k2', json_parse('{}'));
select object(k, v) from bar;

object
------------
{"k1":[1]}
{"k2":{}}
(2 rows)

-- Errors out because DATE type values can't be converted to SUPER type.
select object('k', '2008-12-31'::date);

ERROR:  OBJECT could not convert type date to super
```

## Función OBJECT_TRANSFORM

Transforma un objeto SUPER.

### Sintaxis

```
OBJECT_TRANSFORM(
  input
  [KEEP path1, ...]
  [SET
    path1, value1,
    ...,  ...
  ]
)
```

### Argumentos

input

Expresión que se convierte en un objeto de tipo SUPER.

KEEP

Todos los valores de path especificados en esta cláusula se conservan y se trasladan al objeto de salida.

Esta cláusula es opcional.

path1, path2, ...

Literales de cadena constantes, en formato de componentes de ruta entre comillas dobles delimitados por puntos. Por ejemplo, `'"a"."b"."c"'` es un valor válido. Esto se aplica al parámetro path de las cláusulas KEEP y SET.

SET

Pares de path y value para modificar una ruta existente o añadir una nueva y establecer el valor de esa ruta en el objeto de salida.

Esta cláusula es opcional.

value1, value2, ...

Expresiones que se resuelven en valores de tipo SUPER. Tenga en cuenta que los tipos numérico, de texto y booleano se pueden resolver en SUPER.

Tipo de retorno

`SUPER`

Notas de uso

OBJECT_TRANSFORM devuelve un objeto de tipo SUPER que contiene los valores de ruta input que se han especificado en KEEP y los pares de path y value que se han especificado en SET.

Si tanto KEEP como SET están vacíos, OBJECT_TRANSFORM devuelve input.

Si input no es un object de tipo SUPER, OBJECT_TRANSFORM devuelve input, independientemente de los valores KEEP o SET.

Ejemplo

En el siguiente ejemplo se transforma un objeto SUPER en otro objeto SUPER.

```
CREATE TABLE employees (
    col_person SUPER
```

```
);

INSERT INTO employees
VALUES
    (
        json_parse('
            {
                "name": {
                    "first": "John",
                    "last": "Doe"
                },
                "age": 25,
                "ssn": "111-22-3333",
                "company": "Company Inc.",
                "country": "U.S."
            }
        ')
    ),
    (
        json_parse('
            {
                "name": {
                    "first": "Jane",
                    "last": "Appleseed"
                },
                "age": 34,
                "ssn": "444-55-7777",
                "company": "Organization Org.",
                "country": "Ukraine"
            }
        ')
    )
;

SELECT
    OBJECT_TRANSFORM(
        col_person
        KEEP
            '"name"."first"',
            '"age"',
            '"company"',
            '"country"'
        SET
            '"name"."first"', UPPER(col_person.name.first::TEXT),
```

```
            '"age"', col_person.age + 5,
            '"company"', 'Amazon'
    ) AS col_person_transformed
FROM employees;

--This result is formatted for ease of reading.
                  col_person_transformed
---------------------------------------------------------------
{
    "name": {
        "first": "JOHN"
    },
    "age": 30,
    "company": "Amazon",
    "country": "U.S."
}
{
    "name": {
        "first": "JANE"
    },
    "age": 39,
    "company": "Amazon",
    "country": "Ukraine"
}
```

## Función UPPER_ATTRIBUTE_NAMES

Convierte todos los nombres de los atributos aplicables de un valor SUPER a mayúsculas, con la misma rutina de conversión de mayúsculas y minúsculas que Función UPPER. UPPER_ATTRIBUTE_NAMES admite caracteres multibyte UTF-8 de hasta un máximo de cuatro bytes por carácter.

Para convertir los nombres del atributo SUPER a minúsculas, use Función LOWER_ATTRIBUTE_NAMES.

### Sintaxis

```
UPPER_ATTRIBUTE_NAMES(super_expression)
```

### Argumentos

super_expression

Una expresión SUPER.

### Tipo de retorno

SUPER

### Ejemplos

Conversión de los nombres del atributo SUPER a mayúsculas

En el siguiente ejemplo, se utiliza UPPER_ATTRIBUTE_NAMES para convertir los nombres de los atributos de todos los valores SUPER de una tabla.

```
-- Create a table and insert several SUPER values.
CREATE TABLE t (i INT, s SUPER);

INSERT INTO t VALUES
  (1, NULL),
  (2, 'a'::SUPER),
  (3, JSON_PARSE('{"AttributeName": "b"}')),
  (4, JSON_PARSE(
     '[{"Subobject": {"c": "c"},
       "Subarray": [{"d": "d"}, "e"]
      }]'));

-- Convert all attribute names to uppercase.
UPDATE t SET s = UPPER_ATTRIBUTE_NAMES(s);

SELECT i, s FROM t ORDER BY i;

 i |                         s
---+---------------------------------------------------
 1 | NULL
 2 | "a"
 3 | {"ATTRIBUTENAME":"B"}
 4 | [{"SUBOBJECT":{"C":"c"},"SUBARRAY":[{"D":"d"}, "e"]}]
```

Observe cómo funciona UPPER_ATTRIBUTE_NAMES.

- Los valores NULL y los valores escalares SUPER como `"a"` permanecen inalterados.
- En un objeto SUPER, todos los nombres de los atributos se cambian a mayúsculas, pero los valores de los atributos, como `"b"`, permanecen inalterados.
- UPPER_ATTRIBUTE_NAMES se aplica de forma recursiva a cualquier objeto SUPER que esté anidado dentro de una matriz SUPER o dentro de otro objeto.

Uso de UPPER_ATTRIBUTE_NAMES en un objeto SUPER con nombres de atributos duplicados

Si un objeto SUPER contiene atributos cuyos nombres solo difieren en mayúsculas y minúsculas, UPPER_ATTRIBUTE_NAMES generará un error. Considere el siguiente ejemplo.

```
SELECT UPPER_ATTRIBUTE_NAMES(JSON_PARSE('{"A": "A", "a": "a"}'));

error:   Invalid input
code:    8001
context: SUPER value has duplicate attributes after case conversion.
```

## Funciones espaciales

Las relaciones entre los objetos geométricos se basan en el Modelo de 9 intersecciones dimensionalmente extendido (DE-9IM). Este modelo define predicados tales como igual, contiene y cubre. Para más información acerca de la definición de las relaciones espaciales, vea DE-9IM en Wikipedia.

Para obtener más información sobre cómo utilizar datos espaciales con Amazon Redshift, consulte Consulta de datos espaciales en Amazon Redshift.

Amazon Redshift proporciona funciones espaciales que operan con tipos de datos `GEOMETRY` y `GEOGRAPHY`. A continuación, se enumeran las funciones que son compatibles con el tipo de datos `GEOGRAPHY`:

- ST_Area
- ST_AsEWKT
- ST_AsGeoJSON
- ST_AsHexEWKB
- ST_AsHexWKB
- ST_AsText

- ST_Distance
- ST_GeogFromText
- ST_GeogFromWKB
- ST_Length
- ST_NPoints
- ST_Perimeter

A continuación, se enumera el conjunto completo de funciones espaciales que son compatibles con Amazon Redshift.

Temas

- AddBBox
- DropBBox
- GeometryType
- H3_Boundary
- H3_Center
- H3_FromLongLat
- H3_FromPoint
- H3_Polyfill
- ST_AddPoint
- ST_Angle
- ST_Area
- ST_AsBinary
- ST_AsEWKB
- ST_AsEWKT
- ST_AsGeoJSON
- ST_AsHexWKB
- ST_AsHexEWKB
- ST_AsText
- ST_Azimuth
- ST_Boundary

- ST_Buffer
- ST_Centroid
- ST_Collect
- ST_Contains
- ST_ContainsProperly
- ST_ConvexHull
- ST_CoveredBy
- ST_Covers
- ST_Crosses
- ST_Dimension
- ST_Disjoint
- ST_Distance
- ST_DistanceSphere
- ST_DWithin
- ST_EndPoint
- ST_Envelope
- ST_Equals
- ST_ExteriorRing
- ST_Force2D
- ST_Force3D
- ST_Force3DM
- ST_Force3DZ
- ST_Force4D
- ST_GeoHash
- ST_GeogFromText
- ST_GeogFromWKB
- ST_GeometryN
- ST_GeometryType
- ST_GeomFromEWKB
- ST_GeomFromEWKT

- ST_GeomFromGeoHash
- ST_GeomFromGeoJSON
- ST_GeomFromGeoSquare
- ST_GeomFromText
- ST_GeomFromWKB
- ST_GeoSquare
- ST_InteriorRingN
- ST_Intersects
- ST_Intersection
- ST_IsPolygonCCW
- ST_IsPolygonCW
- ST_IsClosed
- ST_IsCollection
- ST_IsEmpty
- ST_IsRing
- ST_IsSimple
- ST_IsValid
- ST_Length
- ST_LengthSphere
- ST_Length2D
- ST_LineFromMultiPoint
- ST_LineInterpolatePoint
- ST_M
- ST_MakeEnvelope
- ST_MakeLine
- ST_MakePoint
- ST_MakePolygon
- ST_MemSize
- ST_MMax
- ST_MMin

- ST_Multi
- ST_NDims
- ST_NPoints
- ST_NRings
- ST_NumGeometries
- ST_NumInteriorRings
- ST_NumPoints
- ST_Perimeter
- ST_Perimeter2D
- ST_Point
- ST_PointN
- ST_Points
- ST_Polygon
- ST_RemovePoint
- ST_Reverse
- ST_SetPoint
- ST_SetSRID
- ST_Simplify
- ST_SRID
- ST_StartPoint
- ST_Touches
- ST_Transform
- ST_Union
- ST_Within
- ST_X
- ST_XMax
- ST_XMin
- ST_Y
- ST_YMax
- ST_YMin

- ST_Z
- ST_ZMax
- ST_ZMin
- SupportsBBox

## AddBBox

AddBBox devuelve una copia de la geometría de entrada que admite la codificación con un cuadro delimitador calculado previamente. Para obtener más información acerca de la compatibilidad con los cuadros delimitadores, consulte Cuadro delimitador.

### Sintaxis

```
AddBBox(geom)
```

### Argumentos

geom

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

### Tipo de devolución

GEOMETRY

Si geom es nulo, entonces se devuelve el valor nulo.

### Ejemplos

El siguiente SQL devuelve una copia de una geometría de polígono de entrada que admite la codificación con un cuadro delimitador.

```
SELECT ST_AsText(AddBBox(ST_GeomFromText('POLYGON((0 0,1 0,0 1,0 0))')));
```

```
 st_astext
----------
 POLYGON((0 0,1 0,0 1,0 0))
```

## DropBBox

DropBBox devuelve una copia de la geometría de entrada que no admite la codificación con un cuadro delimitador calculado previamente. Para obtener más información acerca de la compatibilidad con los cuadros delimitadores, consulte Cuadro delimitador.

### Sintaxis

```
DropBBox(geom)
```

### Argumentos

geom

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

### Tipo de devolución

GEOMETRY

Si geom es nulo, entonces se devuelve el valor nulo.

### Ejemplos

El siguiente SQL devuelve una copia de una geometría de polígono de entrada que no admite la codificación con un cuadro delimitador.

```
SELECT ST_AsText(DropBBox(ST_GeomFromText('POLYGON((0 0,1 0,0 1,0 0))')));
```

```
 st_astext
----------
 POLYGON((0 0,1 0,0 1,0 0))
```

## GeometryType

GeometryType devuelve el subtipo de una geometría de entrada de datos como una cadena.

### Sintaxis

```
GeometryType(geom)
```

Argumentos

geom

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

Tipo de devolución

VARCHAR representa el subtipo de geom.

Si geom es nulo, entonces se devuelve el valor nulo.

Los valores retornados son los siguientes.

| Valor de cadena devuelto para geometrías 2D, 3DZ y 4D | Valor de cadena devuelto para geometrías 3DM | Subtipo de geometría. |
|---|---|---|
| POINT | POINTM | Se devuelve si geom es un subtipo de POINT |
| LINESTRING | LINESTRINGM | Se devuelve si geom es un subtipo de LINESTRING |
| POLYGON | POLYGONM | Se devuelve si geom es un subtipo de POLYGON |
| MULTIPOINT | MULTIPOINTM | Se devuelve si geom es un subtipo de MULTIPOINT |
| MULTILINESTRING | MULTILINESTRINGM | Se devuelve si geom es un subtipo de MULTILINESTRING |
| MULTIPOLYGON | MULTIPOLYGONM | Se devuelve si geom es un subtipo de MULTIPOLYGON |
| GEOMETRYCOLLECTION | GEOMETRYCOLLECTIONM | Se devuelve si geom es un subtipo de GEOMETRYCOLLECTION |

### Ejemplos

El siguiente SQL convierte una representación en texto conocido (WKT) de un polígono y devuelve el subtipo de GEOMETRY como una cadena.

```
SELECT GeometryType(ST_GeomFromText('POLYGON((0 2,1 1,0 -1,0 2))'));
```

```
geometrytype
-------------
 POLYGON
```

## H3_Boundary

H3_Boundary devuelve el límite de un ID de celda H3 a partir de un índice de entrada. Para obtener más información acerca de la indexación H3, consulte H3.

### Sintaxis

```
H3_Boundary(index)
```

### Argumentos

índice

Un valor de tipo de datos BIGINT o VARCHAR que representa el índice de una celda H3. O una expresión que dé como resultado uno de estos tipos de datos.

### Tipo de devolución

POLYGON: representa el polígono con el identificador del sistema de referencia espacial (SRID) de 0.

Si index no es válido, se devuelve un error.

### Ejemplos

El siguiente código SQL introduce un VARCHAR que representa el índice de una celda H3 y devuelve un POLYGON con SRID 0 que representa el límite de la celda H3 de entrada. La salida de H3_Boundary es la entrada de ST_AwEWKT para mostrarla en representación de texto ampliado conocido (EWKT).

```
SELECT ST_AsEWKT(H3_Boundary('8025fffffffffff'));
```

```
 st_asewkt
-----------------------------------------------------------------------------------------------------------
 POLYGON((87.7729641223953 52.6542030078627,86.7082098104731
 40.3127667561966,98.2285042557705 33.6210697806835,110.694610548823
 37.163896485796,116.212895637138 47.3094513028131,106.40349563788
 56.210610737585,87.7729641223953 52.6542030078627))
```

El siguiente código SQL introduce un BIGINT que representa el índice de una celda H3 y devuelve un POLYGON con SRID 0 que representa el límite de la celda H3 de entrada. La salida de H3_Boundary es la entrada de ST_AwEWKT para mostrarla en representación de texto ampliado conocido (EWKT).

```
SELECT ST_AsEWKT(H3_Boundary(577129255373111295));
```

```
 st_asewkt
-----------------------------------------------------------------------------------------------------------
POLYGON((87.7729641223953 52.6542030078627,86.7082098104731
 40.3127667561966,98.2285042557705 33.6210697806835,110.694610548823
 37.163896485796,116.212895637138 47.3094513028131,106.40349563788
 56.210610737585,87.7729641223953 52.6542030078627))
```

## H3_Center

H3_Center devuelve el centroide de un ID de celda H3 a partir de un índice de entrada. Para obtener más información acerca de la indexación H3, consulte H3.

### Sintaxis

```
H3_Center(index)
```

### Argumentos

índice

Un valor de tipo de datos BIGINT o VARCHAR que representa el índice de una celda H3. O una expresión que dé como resultado uno de estos tipos de datos.

Tipo de devolución

`POINT`: representa el centroide de la celda H3 con el identificador del sistema de referencia espacial (SRID) de `0`.

Si index no es válido, se devuelve un error.

Ejemplos

El siguiente código SQL introduce un `VARCHAR` que representa el índice de una celda H3 y devuelve un POINT con SRID de 0 que representa el centroide de la celda H3 de entrada.

```
SELECT H3_Center('8025fffffffffff');
```

```
 h3_center
------------------------------------------
 010100000070707A550B605940AEE9D70B327E4640
```

El siguiente código SQL introduce un `BIGINT` que representa el índice de una celda H3 y devuelve un POINT con SRID 0 que representa el centroide de la celda H3 de entrada.

```
SELECT H3_Center(577129255373111295);
```

```
 h3_center
------------------------------------------
 010100000070707A550B605940AEE9D70B327E4640
```

El siguiente código SQL introduce un `VARCHAR` que representa el índice de una celda H3 y devuelve un POINT con SRID 0 que representa el centroide de la celda H3 de entrada. La salida de H3_Center es la entrada de ST_AwEWKT para mostrarla en representación de texto ampliado conocido (EWKT).

```
SELECT ST_AsEWKT(H3_Center('8075fffffffffff'));
```

```
 st_asewkt
```

```
----------------------------------------
POINT(-5.24539029677733 2.30088211162675)
```

## H3_FromLongLat

H3_FromLongLat devuelve el ID de celda H3 correspondiente a partir de una entrada de longitud, latitud y resolución. Para obtener más información acerca de la indexación H3, consulte H3.

### Sintaxis

```
H3_FromLongLat(longitude, lattitude, resolution)
```

### Argumentos

longitude

Un valor de tipo de datos `DOUBLE PRECISION` o una expresión que toma el valor de un tipo `DOUBLE PRECISION`.

latitude

Un valor de tipo de datos `DOUBLE PRECISION` o una expresión que toma el valor de un tipo `DOUBLE PRECISION`.

resolution

Es un valor de tipo de dato `INTEGER` o una expresión que toma el valor de un tipo `INTEGER`. El valor representa la resolución del sistema de cuadrícula H3. El valor debe ser un entero entre 0 y 15, ambos incluidos. `0` es el más basto y `15` el más fino.

### Tipo de devolución

`BIGINT`: representa el ID de celda H3.

Si resolution está fuera de los límites, se devuelve un error.

### Ejemplos

La siguiente SQL devuelve el ID de celda H3 en términos de longitud `0`, latitud `0` y resolución `10`.

```
SELECT H3_FromLongLat(0, 0, 10);
```

```
 h3_fromlonglat
-------------------
 623560421467684863
```

## H3_FromPoint

H3_FromPoint devuelve el ID de celda H3 correspondiente a partir de un punto y resolución de geometría de entrada. Para obtener más información acerca de la indexación H3, consulte H3.

### Sintaxis

```
H3_FromPoint(geom, resolution)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`. geom debe ser un `POINT`.

resolution

Es un valor de tipo de dato `INTEGER` o una expresión que toma el valor de un tipo `INTEGER`. El valor representa la resolución del sistema de cuadrícula H3. El valor debe ser un entero entre 0 y 15, ambos incluidos. `0` es el más basto y `15` el más fino.

### Tipo de devolución

`BIGINT`: representa el ID de celda H3.

Si geom no es `POINT`, se devuelve un error.

Si resolution está fuera de los límites, se devuelve un error.

Si geom está vacío, entonces se devuelve el valor NULL.

### Ejemplos

La siguiente SQL devuelve el ID de celda H3 en términos de punto `0,0` y resolución `10`.

```
SELECT H3_FromPoint(ST_GeomFromText('POINT(0 0)'), 10);
```

```
 h3_frompoint
-------------------
 623560421467684863
```

## H3_Polyfill

H3_Polyfill devuelve los ID de celda H3 correspondientes a los hexágonos y pentágonos contenidos incluidos en el polígono de entrada de la resolución dada. Para obtener más información acerca de la indexación H3, consulte H3.

### Sintaxis

```
H3_Polyfill(geom, resolution)
```

### Argumentos

geom

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY. geom debe ser un POLYGON.

resolution

Es un valor de tipo de dato INTEGER o una expresión que toma el valor de un tipo INTEGER. El valor representa la resolución del sistema de cuadrícula H3. El valor debe ser un entero entre 0 y 15, ambos incluidos. 0 es el más basto y 15 el más fino.

### Tipo de devolución

SUPER: representa una lista de ID de celda H3.

Si geom no es POLYGON, se devuelve un error.

Si resolution está fuera de los límites, se devuelve un error.

Si geom está vacío, entonces se devuelve el valor NULL.

Ejemplos

La siguiente SQL devuelve una matriz de tipos de datos SUPER con los ID de celda H3 de un polígono y la resolución 4.

```
SELECT H3_Polyfill(ST_GeomFromText('POLYGON((0 0, 0 1, 1 1, 1 0, 0 0))'), 4);
```

```
 h3_polyfill
-------------------------------------------------------------------------------------------------
 [596538848238895103,596538805289222143,596538856828829695,596538813879156735,5965379205259591
```

## ST_AddPoint

ST_AddPoint devuelve una geometría de linestring que es la misma que la geometría de entrada con un punto agregado. Si se proporciona un índice, entonces el punto se agrega en la posición del índice. Si el índice es -1 o no se proporciona, entonces el punto se anexa a la linestring.

El índice tiene base cero. El identificador del sistema de referencia espacial (SRID) del resultado es el mismo que el de la geometría de entrada.

La dimensión de la geometría devuelta es la misma que la del valor geom1. Si geom1 y geom2 tienen diferentes dimensiones, geom2 se proyecta a la dimensión de geom1.

Sintaxis

```
ST_AddPoint(geom1, geom2)
```

```
ST_AddPoint(geom1, geom2, index)
```

Argumentos

geom1

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`. El subtipo debe ser `LINESTRING` .

geom2

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`. El subtipo debe ser `POINT` . El punto puede ser el punto vacío.

índice

Un valor del tipo de datos INTEGER que representa la posición de un índice con base cero.

Tipo de devolución

GEOMETRY

Si geom1, geom2, o índice es null, se devuelve el valor null.

Si geom2 es el punto vacío, se devuelve una copia de geom1.

Si geom1 no es LINESTRING, se devuelve un error.

Si geom2 no es POINT, se devuelve un error.

Si el índice está fuera de rango, se devuelve un error. Los valores válidos para la posición de índice son -1 o un valor entre 0 y ST_NumPoints(geom1).

Ejemplos

El siguiente SQL agrega un punto a una linestring para convertirla en una linestring cerrada.

```
WITH tmp(g) AS (SELECT ST_GeomFromText('LINESTRING(0 0,10 0,10 10,5 5,0 5)',4326))
SELECT ST_AsEWKT(ST_AddPoint(g, ST_StartPoint(g))) FROM tmp;
```

```
 st_asewkt
-------------------------------------------------
 SRID=4326;LINESTRING(0 0,10 0,10 10,5 5,0 5,0 0)
```

El siguiente SQL agrega un punto a una posición específica en una linestring.

```
WITH tmp(g) AS (SELECT ST_GeomFromText('LINESTRING(0 0,10 0,10 10,5 5,0 5)',4326))
SELECT ST_AsEWKT(ST_AddPoint(g, ST_SetSRID(ST_Point(5, 10), 4326), 3)) FROM tmp;
```

```
 st_asewkt
-------------------------------------------------
 SRID=4326;LINESTRING(0 0,10 0,10 10,5 10,5 5,0 5)
```

## ST_Angle

ST_Angle devuelve el ángulo en radianes entre los puntos medidos en el sentido de las agujas del reloj de la siguiente manera:

- Si hay tres puntos de entrada, entonces el ángulo devuelto P1-P2-P3 se mide como si el ángulo se obtuviera al rotar de P1 a P3 alrededor de P2 en el sentido de las agujas del reloj.
- Si hay cuatro puntos de entrada, se devuelve el ángulo devuelto en el sentido de las agujas del reloj formado por las líneas dirigidas P1-P2 y P3-P4. Si la entrada es un caso degenerado (es decir, P1 es igual a P2 o P3 es igual a P4), entonces se devuelve el valor null.

El valor de retorno está en radianes y en el rango [0, 2π).

ST_Angle opera en proyecciones 2D de las geometrías de entrada.

Sintaxis

```
ST_Angle(geom1, geom2, geom3)
```

```
ST_Angle(geom1, geom2, geom3, geom4)
```

Argumentos

geom1

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`. El subtipo debe ser `POINT` .

geom2

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`. El subtipo debe ser `POINT` .

geom3

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`. El subtipo debe ser `POINT` .

geom4

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`. El subtipo debe ser `POINT` .

Tipo de devolución

`DOUBLE PRECISION`.

Si geom1 es igual a geom2 o geom2 es igual a geom3, se devuelve un valor nulo.

Si geom1, geom2, geom3 o geom4 es null, se devuelve un valor nulo.

Si alguna de las funciones geom1, geom2, geom3 o geom4 es el punto vacío, se devuelve un error.

Si geom1, geom2, geom3 y geom4 no tienen el mismo valor en el identificador de sistema de referencia espacial (SRID), se devuelve un error.

Ejemplos

El siguiente SQL devuelve el ángulo convertido en grados de tres puntos de entrada.

```
SELECT ST_Angle(ST_Point(1,1), ST_Point(0,0), ST_Point(1,0)) / Pi() * 180.0 AS angle;
```

```
 angle
---------------
    45
```

El siguiente SQL devuelve el ángulo convertido en grados de cuatro puntos de entrada.

```
SELECT ST_Angle(ST_Point(1,1), ST_Point(0,0), ST_Point(1,0), ST_Point(2,0)) / Pi() *
 180.0 AS angle;
```

```
 angle
---------------
   225
```

## ST_Area

Con respecto a una geometría de entrada, ST_Area devuelve el área cartesiana de una proyección 2D. Las unidades de área son las mismas que las unidades en las que se expresan las coordenadas de la geometría de entrada. Para puntos, cadenas de línea, multipuntos y linestring múltiples, la función devuelve 0. Para las colecciones de geometría, devuelve la suma de las áreas de las geometrías de la colección.

Con respecto a una geografía de entrada, ST_Area devuelve el área geodésica de la proyección 2D de una geografía de área de entrada calculada sobre el esferoide que determina el SRID. La unidad de longitud se expresa en metros cuadrados. La función devuelve un valor cero (0) para las geografías de puntos, multipuntos y lineales. Cuando la entrada es una colección de geometría, la función devuelve la suma de las áreas de las geografías de la colección.

Sintaxis

```
ST_Area(geo)
```

Argumentos

geo

Un valor de tipo de datos GEOMETRY o GEOGRAPHY, o una expresión que toma el valor de un tipo GEOMETRY o GEOGRAPHY.

Tipo de devolución

DOUBLE PRECISION

Si geom es nulo, entonces se devuelve el valor nulo.

Ejemplos

El siguiente SQL devuelve el área cartesiana de un multipolygon.

```
SELECT ST_Area(ST_GeomFromText('MULTIPOLYGON(((0 0,10 0,0 10,0 0)),((10 0,20 0,20 10,10
 0)))'));
```

```
 st_area
---------
     100
```

El siguiente SQL devuelve el área de un polígono en una geografía.

```
SELECT ST_Area(ST_GeogFromText('polygon((34 35, 28 30, 25 34, 34 35))'));
```

```
     st_area
------------------
 201824655743.383
```

El siguiente SQL devuelve un valor cero para una geografía lineal.

```
SELECT ST_Area(ST_GeogFromText('multipoint(0 0, 1 1, -21.32 121.2)'));
```

```
 st_area
---------
       0
```

## ST_AsBinary

ST_AsBinary devuelve la representación binaria conocida (WKB) hexadecimal de una geometría de entrada. En el caso de las geometrías 3DZ, 3DM y 4D, ST_AsBinary utiliza el valor estándar de Open Geospatial Consortium (OGC) para el tipo de geometría.

Sintaxis

```
ST_AsBinary(geom)
```

Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

Tipo de devolución

`VARBYTE`

Si geom es nulo, entonces se devuelve el valor nulo.

Ejemplos

El siguiente SQL devuelve la representación WKB hexadecimal de un polígono.

```
SELECT ST_AsBinary(ST_GeomFromText('POLYGON((0 0,0 1,1 1,1 0,0 0))',4326));
```

```
st_asbinary
--------------------------------
0103000000010000000500000000000000000000000000000000000000000000000000F03F00000
```

## ST_AsEWKB

ST_AsEWKT devuelve la representación binaria conocida extendida (EWKT) de una geometría de datos de entrada. En el caso de las geometrías 3DZ, 3DM y 4D, ST_AsEWKB utiliza el valor estándar de Open Geospatial Consortium (OGC) para el tipo de geometría.

### Sintaxis

```
ST_AsEWKB(geom)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

### Tipo de devolución

`VARBYTE`

Si geom es nulo, entonces se devuelve el valor nulo.

### Ejemplos

El siguiente SQL devuelve la representación EWKB hexadecimal de un polígono.

```
SELECT ST_AsEWKB(ST_GeomFromText('POLYGON((0 0,0 1,1 1,1 0,0 0))',4326));
```

```
st_asewkb
--------------------------------
0103000020E6100000010000000500000000000000000000000000000000000000000000000000F
```

## ST_AsEWKT

ST_AsEWKT devuelve la representación de texto conocido extendido (EWKT) de una geometría o una geografía de entrada de datos. Para las geometrías 3DZ, 3DM y 4D, ST_AsEWKT agrega Z, M o ZM al valor WKT del tipo de geometría.

### Sintaxis

```
ST_AsEWKT(geo)
```

```
ST_AsEWKT(geo, precision)
```

### Argumentos

geo

Un valor de tipo de datos GEOMETRY o GEOGRAPHY, o una expresión que toma el valor de un tipo GEOMETRY o GEOGRAPHY.

precisión

Un valor de tipo de datos INTEGER. En el caso de las geometrías, las coordenadas de geo se muestran mediante la precisión especificada de 1 a 20. Si no se especifica precision, el valor predeterminado es 15. En el caso de las geometrías, las coordenadas de geo se muestran mediante la precisión especificada. Si no se especifica precision, el valor predeterminado es 15.

### Tipo de devolución

VARCHAR

Si geo es nulo, entonces se devuelve el valor nulo.

Si precision es null, entonces se devuelve el valor null.

Si el resultado es mayor de 64KB de VARCHAR, entonces se devuelve un error.

### Ejemplos

El siguiente SQL devuelve la representación EWKT de una linestring.

```
SELECT ST_AsEWKT(ST_GeomFromText('LINESTRING(3.141592653589793
 -6.283185307179586,2.718281828459045 -1.414213562373095)', 4326));
```

```
st_asewkt
-------------------------------
 SRID=4326;LINESTRING(3.14159265358979 -6.28318530717959,2.71828182845905
 -1.41421356237309)
```

El siguiente SQL devuelve la representación EWKT de una linestring. Las coordenadas de la geometría se muestran con una precisión de seis dígitos.

```
SELECT ST_AsEWKT(ST_GeomFromText('LINESTRING(3.141592653589793
 -6.283185307179586,2.718281828459045 -1.414213562373095)', 4326), 6);
```

```
st_asewkt
-------------------------------
 SRID=4326;LINESTRING(3.14159 -6.28319,2.71828 -1.41421)
```

El siguiente SQL devuelve la representación EWKT de una geografía.

```
SELECT ST_AsEWKT(ST_GeogFromText('LINESTRING(110 40, 2 3, -10 80, -7 9)'));
```

```
                 st_asewkt
-------------------------------------------
 SRID=4326;LINESTRING(110 40,2 3,-10 80,-7 9)
```

## ST_AsGeoJSON

ST_AsGeoJSON devuelve la representación GeoJSON de una geometría o una geografía de entrada de datos. Para saber más acerca de GeoJSON, vea GeoJSON en Wikipedia.

En el caso de las geometrías 3DZ y 4D, la geometría de salida es una proyección 3DZ de la geometría 3DZ o 4D de entrada. Es decir, las coordenadas x, y y z están presentes en la salida. En el caso de las geometrías 3DM, la geometría de salida es una proyección 2D de la geometría 3DM de entrada. Es decir, solo las coordenadas x and y están presentes en la salida.

En el caso de las geografías de entrada de datos, ST_AsGeoJSON devuelve la representación GeoJSON de una geografía de entrada de datos. Las coordenadas de la geografía se muestran mediante la precisión especificada.

### Sintaxis

```
ST_AsGeoJSON(geo)
```

```
ST_AsGeoJSON(geo, precision)
```

### Argumentos

geo

Un valor de tipo de datos GEOMETRY o GEOGRAPHY, o una expresión que toma el valor de un tipo GEOMETRY o GEOGRAPHY.

precisión

Un valor de tipo de datos INTEGER. En el caso de las geometrías, las coordenadas de geo se muestran mediante la precisión especificada de 1 a 20. Si no se especifica precision, el valor predeterminado es 15. En el caso de las geometrías, las coordenadas de geo se muestran mediante la precisión especificada. Si no se especifica precision, el valor predeterminado es 15.

### Tipo de devolución

VARCHAR

Si geo es nulo, entonces se devuelve el valor nulo.

Si precision es null, entonces se devuelve el valor null.

Si el resultado es mayor de 64KB de VARCHAR, entonces se devuelve un error.

### Ejemplos

El siguiente SQL devuelve la representación GeoJSON de una linestring.

```
SELECT ST_AsGeoJSON(ST_GeomFromText('LINESTRING(3.141592653589793
 -6.283185307179586,2.718281828459045 -1.414213562373095)'));
```

```
st_asgeojson
--------------------------------------------------------------------------------
```

```
 {"type":"LineString","coordinates":[[3.14159265358979,-6.28318530717959],
[2.71828182845905,-1.4142135623709]]}
```

El siguiente SQL devuelve la representación GeoJSON de una linestring. Las coordenadas de la geometría se muestran con una precisión de seis dígitos.

```
SELECT ST_AsGeoJSON(ST_GeomFromText('LINESTRING(3.141592653589793
 -6.283185307179586,2.718281828459045 -1.414213562373095)'), 6);
```

```
st_asgeojson
--------------------------------------------------------------------------------
  {"type":"LineString","coordinates":[[3.14159,-6.28319],[2.71828,-1.41421]]}
```

El siguiente SQL devuelve la representación GeoJSON de una geografía.

```
SELECT ST_AsGeoJSON(ST_GeogFromText('LINESTRING(110 40, 2 3, -10 80, -7 9)'));
```

```
                            st_asgeojson
-----------------------------------------------------------------------
 {"type":"LineString","coordinates":[[110,40],[2,3],[-10,80],[-7,9]]}
```

## ST_AsHexWKB

ST_AsHexWKB devuelve la representación binaria conocida (WKB) hexadecimal de una geometría o una geografía de entrada de datos mediante caracteres ASCII hexadecimales (de 0 a 9, de A a F). En el caso de geometrías o geografías 3DZ, 3DM y 4D, ST_AsHexWKB, utiliza el valor estándar del Open Geospatial Consortium (OGC) correspondiente al tipo de geometría o geografía.

Sintaxis

```
ST_AsHexWKB(geo)
```

Argumentos

geo

Un valor de tipo de datos GEOMETRY o GEOGRAPHY, o una expresión que toma el valor de un tipo GEOMETRY o GEOGRAPHY.

Tipo de devolución

VARCHAR

Si geo es nulo, entonces se devuelve el valor nulo.

Si el resultado es mayor de 64KB de VARCHAR, entonces se devuelve un error.

Ejemplos

El siguiente SQL devuelve la representación WKB hexadecimal de un polígono en una geometría.

```
SELECT ST_AsHexWKB(ST_GeomFromText('POLYGON((0 0,0 1,1 1,1 0,0 0))',4326));
```

```
st_ashexwkb
--------------------------------
0103000000010000000500000000000000000000000000000000000000000000000000F03F00000
```

El siguiente SQL devuelve la representación WKB hexadecimal de un polígono en una geografía.

```
SELECT ST_AsHexWKB(ST_GeogFromText('POLYGON((0 0,0 1,1 1,1 0,0 0))'));
```

```
st_ashexwkb
--------------------------------
0103000000010000000500000000000000000000000000000000000000000000000000F03F00000
```

## ST_AsHexEWKB

ST_AsHexEWKB devuelve la representación binaria conocida extendida (EWKB) de una geometría o una geografía de entrada de datos mediante caracteres ASCII hexadecimales (de 0 a 9, de A a F). En el caso de geometrías o geografías 3DZ, 3DM y 4D, ST_AsHexEWKB utiliza el valor WKB extendido PostGIS correspondiente al tipo de geometría o geografía.

Sintaxis

```
ST_AsHexEWKB(geo)
```

### Argumentos

geo

Un valor de tipo de datos GEOMETRY o GEOGRAPHY, o una expresión que toma el valor de un tipo GEOMETRY o GEOGRAPHY.

### Tipo de devolución

VARCHAR

Si geo es nulo, entonces se devuelve el valor nulo.

Si el resultado es mayor de 64KB de VARCHAR, entonces se devuelve un error.

### Ejemplos

El siguiente SQL devuelve la representación EWKB hexadecimal de un polígono en una geometría.

```
SELECT ST_AsHexEWKB(ST_GeomFromText('POLYGON((0 0,0 1,1 1,1 0,0 0))',4326));
```

```
st_ashexewkb
--------------------------------
0103000020E6100000010000000500000000000000000000000000000000000000000000000000F0
```

El siguiente SQL devuelve la representación EWKB hexadecimal de un polígono en una geografía.

```
SELECT ST_AsHexEWKB(ST_GeogFromText('POLYGON((0 0,0 1,1 1,1 0,0 0))'));
```

```
st_ashexewkb
--------------------------------
0103000020E6100000010000000500000000000000000000000000000000000000000000000000F0
```

## ST_AsText

ST_AsText devuelve la representación de texto conocido (WKT) de una geometría o una geografía de datos de entrada. En el caso de geometrías o geografías 3DZ, 3DM y 4D, ST_AsEWKT agrega Z, M o ZM al valor WKT correspondiente al tipo de geometría o geografía.

### Sintaxis

```
ST_AsText(geo)
```

```
ST_AsText(geo, precision)
```

### Argumentos

geo

Un valor de tipo de datos GEOMETRY o GEOGRAPHY, o una expresión que toma el valor de un tipo GEOMETRY o GEOGRAPHY.

precisión

Un valor de tipo de datos INTEGER. En el caso de las geometrías, las coordenadas de geo se muestran mediante la precisión especificada de 1 a 20. Si no se especifica precision, el valor predeterminado es 15. En el caso de las geografías, las coordenadas de geo se muestran mediante la precisión especificada. Si no se especifica precision, el valor predeterminado es 15.

### Tipo de devolución

VARCHAR

Si geo es nulo, entonces se devuelve el valor nulo.

Si precision es null, entonces se devuelve el valor null.

Si el resultado es mayor de 64KB de VARCHAR, entonces se devuelve un error.

### Ejemplos

El siguiente SQL devuelve la representación WKT de una linestring.

```
SELECT ST_AsText(ST_GeomFromText('LINESTRING(3.141592653589793
 -6.283185307179586,2.718281828459045 -1.414213562373095)', 4326));
```

```
st_astext
-------------------------------
LINESTRING(3.14159265358979 -6.28318530717959,2.71828182845905 -1.41421356237309)
```

El siguiente SQL devuelve la representación WKT de una linestring. Las coordenadas de la geometría se muestran con una precisión de seis dígitos.

```
SELECT ST_AsText(ST_GeomFromText('LINESTRING(3.141592653589793
 -6.283185307179586,2.718281828459045 -1.414213562373095)', 4326), 6);
```

```
st_astext
---------------------------------------------
 LINESTRING(3.14159 -6.28319,2.71828 -1.41421)
```

El siguiente SQL devuelve la representación WKT de una geografía.

```
SELECT ST_AsText(ST_GeogFromText('LINESTRING(110 40, 2 3, -10 80, -7 9)'));
```

```
             st_astext
-----------------------------------
 LINESTRING(110 40,2 3,-10 80,-7 9)
```

## ST_Azimuth

ST_Azimuth devuelve el acimut cartesiano con base en el norte mediante proyecciones 2D de los dos puntos de entrada.

Sintaxis

```
ST_Azimuth(point1, point2)
```

Argumentos

point1

Un valor `POINT` de tipo de datos `GEOMETRY`. El identificador de sistema de referencia espacial (SRID) de point1 debe coincidir con el SRID de point2.

point2

Un valor `POINT` de tipo de datos `GEOMETRY`. El SRID de point2 debe coincidir con el SRID de point1.

Tipo de devolución

Un número que es un ángulo en radianes de tipo de datos `DOUBLE PRECISION`. Rango de valores de 0 (incluido) hasta 2 pi (no incluido).

Si point1 o point2 son el punto vacío, se devuelve un error.

Si el valor de point1 o point2 es null, entonces se devuelve null.

Si point1 y point2 son iguales, se devuelve null.

Si point1 o point2 no son un punto, se devuelve un error.

Si point1 y point2 no tienen un valor en el identificador de sistema de referencia espacial (SRID), se devuelve un error.

Ejemplos

El siguiente SQL devuelve el acimut de los puntos de entrada de datos.

```
SELECT ST_Azimuth(ST_Point(1,2), ST_Point(5,6));
```

```
st_azimuth
-------------------
 0.7853981633974483
```

## ST_Boundary

ST_Boundary devuelve el límite de una geometría de entrada de la siguiente manera:

- Si la geometría de entrada está vacía (es decir, no contiene puntos), se devuelve tal cual.
- Si la geometría de entrada es un punto o un multipunto no vacío, se devuelve una colección de geometría vacía.
- Si la entrada es una cadena de línea o una cadena de varias líneas, se devuelve un multipunto que contiene todos los puntos del límite. El multipunto puede estar vacío.
- Si la entrada es un polígono que no tiene anillos interiores, se devuelve una cadena de línea cerrada que representa su límite.

- Si la entrada es un polígono que tiene anillos interiores o es un multipolígono, se devuelve una cadena de varias líneas. La cadena de varias líneas contiene todos los límites de todos los anillos de la geometría de área como cadenas de línea cerradas.

Para determinar la igualdad de los puntos, ST_Boundary opera en la proyección 2D de la geometría de entrada. Si la geometría de entrada está vacía, se devuelve una copia de ella en la misma dimensión que la entrada. En el caso de las geometrías 3DM y 4D no vacías, se eliminan sus coordenadas `m`. En el caso especial de las cadenas de varias líneas 3DZ y 4D, las coordenadas z de los puntos de límite de la cadena de varias líneas se calculan como promedios de los distintos valores z de los puntos de límite de la cadena de líneas con la misma proyección 2D.

Sintaxis

```
ST_Boundary(geom)
```

Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

Tipo de devolución

`GEOMETRY`

Si geom es nulo, entonces se devuelve el valor nulo.

Si geom es `GEOMETRYCOLLECTION`, se devuelve un error.

Ejemplos

El siguiente SQL devuelve el límite del polígono de entrada como una cadena de varias líneas.

```
SELECT ST_AsEWKT(ST_Boundary(ST_GeomFromText('POLYGON((0 0,10 0,10 10,0 10,0 0),(1 1,1
 2,2 1,1 1))')));
```

```
st_asewkt
--------------------
 MULTILINESTRING((0 0,10 0,10 10,0 10,0 0),(1 1,1 2,2 1,1 1))
```

## ST_Buffer

ST_Buffer devuelve una geometría 2D que representa todos los puntos cuya distancia desde la geometría de entrada proyectada en el plano cartesiano XY es menor o igual a la distancia de entrada.

### Sintaxis

```
ST_Buffer(geom, distance)
```

```
ST_Buffer(geom, distance, number_of_segments_per_quarter_circle)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

distance

Un valor de tipo de datos `DOUBLE PRECISION` que representa la distancia (o el radio) del búfer.

number_of_segments_per_quarter_circle

Un valor de tipo de datos `INTEGER`. Este valor determina el número de puntos que se aproximan a un cuarto de círculo en torno a cada vértice de la geometría de entrada. Los valores negativos son cero de forma predeterminada. El valor predeterminado es 8.

### Tipo de devolución

`GEOMETRY`

La función ST_Buffer devuelve geometría bidimensional (2D) en el plano cartesiano XY.

Si geom es `GEOMETRYCOLLECTION`, se devuelve un error.

### Ejemplos

El siguiente SQL devuelve el búfer de la cadena de líneas de entrada.

```
SELECT ST_AsEwkt(ST_Buffer(ST_GeomFromText('LINESTRING(1 2,5 2,5 8)'), 2));
```

```
               st_asewkt
POLYGON((-1 2,-0.96157056080646 2.39018064403226,-0.847759065022573
 2.76536686473018,-0.662939224605089 3.11114046603921,-0.414213562373093
 3.4142135623731,-0.111140466039201 3.66293922460509,0.234633135269824
 3.84775906502257,0.609819355967748 3.96157056080646,1 4,3 4,3 8,3.03842943919354
 8.39018064403226,3.15224093497743 8.76536686473018,3.33706077539491
 9.11114046603921,3.58578643762691 9.4142135623731,3.8888595339608
 9.66293922460509,4.23463313526982 9.84775906502257,4.60981935596775
 9.96157056080646,5 10,5.39018064403226 9.96157056080646,5.76536686473018
 9.84775906502257,6.11114046603921 9.66293922460509,6.4142135623731
 9.41421356237309,6.66293922460509 9.1111404660392,6.84775906502258
 8.76536686473017,6.96157056080646 8.39018064403225,7 8,7 2,6.96157056080646
 1.60981935596774,6.84775906502257 1.23463313526982,6.66293922460509
 0.888859533960796,6.41421356237309 0.585786437626905,6.1111404660392
 0.33706077539491,5.76536686473018 0.152240934977427,5.39018064403226
 0.0384294391935391,5 0,1 0,0.609819355967744 0.0384294391935391,0.234633135269821
 0.152240934977427,-0.111140466039204 0.337060775394909,-0.414213562373095
 0.585786437626905,-0.662939224605091 0.888859533960796,-0.847759065022574
 1.23463313526982,-0.961570560806461 1.60981935596774,-1 2))
```

El siguiente SQL devuelve el búfer de la geometría de puntos de entrada que se aproxima a un círculo. Dado que el comando no especifica el número de segmentos por cuarto de círculo, la función utiliza el valor predeterminado de ocho segmentos para aproximarse al cuarto de círculo.

```
SELECT ST_AsEwkt(ST_Buffer(ST_GeomFromText('POINT(3 4)'), 2));
```

```
               st_asewkt
POLYGON((1 4,1.03842943919354 4.39018064403226,1.15224093497743
 4.76536686473018,1.33706077539491 5.11114046603921,1.58578643762691
 5.4142135623731,1.8888595339608 5.66293922460509,2.23463313526982
 5.84775906502257,2.60981935596775 5.96157056080646,3 6,3.39018064403226
 5.96157056080646,3.76536686473019 5.84775906502257,4.11114046603921
 5.66293922460509,4.4142135623731 5.41421356237309,4.66293922460509
 5.1111404660392,4.84775906502258 4.76536686473017,4.96157056080646 4.39018064403225,5
 4,4.96157056080646 3.60981935596774,4.84775906502257 3.23463313526982,4.66293922460509
 2.8888595339608,4.41421356237309 2.58578643762691,4.1111404660392
 2.33706077539491,3.76536686473018 2.15224093497743,3.39018064403226 2.03842943919354,3
 2,2.60981935596774 2.03842943919354,2.23463313526982 2.15224093497743,1.8888595339608
 2.33706077539491,1.58578643762691 2.58578643762691,1.33706077539491
 2.8888595339608,1.15224093497743 3.23463313526982,1.03842943919354 3.60981935596774,1
 4))
```

El siguiente SQL devuelve el búfer de la geometría de puntos de entrada que se aproxima a un círculo. Dado que el comando especifica 3 como el número de segmentos por cuarto de círculo, la función utiliza tres segmentos para aproximarse al cuarto de círculo.

```
SELECT ST_AsEwkt(ST_Buffer(ST_GeomFromText('POINT(3 4)'), 2, 3));
```

```
               st_asewkt
POLYGON((1 4,1.26794919243112 5,2 5.73205080756888,3 6,4
 5.73205080756888,4.73205080756888 5,5 4,4.73205080756888 3,4 2.26794919243112,3 2,2
 2.26794919243112,1.26794919243112 3,1 4))
```

## ST_Centroid

ST_Centroid devuelve un punto que representa un centroide de una geometría, tal y como se indica a continuación:

- En el caso de las geometrías `POINT`, se devuelve el punto cuyas coordenadas constituyen la media de las coordenadas de los puntos de la geometría.
- En el caso de las geometrías `LINESTRING`, se devuelve el punto cuyas coordenadas constituyen la media ponderada de los puntos medios de los segmentos de la geometría, donde las ponderaciones corresponden a las longitudes de esos segmentos.
- En el caso de las geometrías `POLYGON`, se devuelve el punto cuyas coordenadas constituyen la media ponderada de los centroides de una triangulación de la geometría de área en la que las ponderaciones corresponden a las áreas de los triángulos que componen dicha triangulación.
- En el caso de las colecciones de geometrías, se devuelve la media ponderada de los centroides de las geometrías que tienen una dimensión topológica máxima dentro de la colección de geometrías.

### Sintaxis

```
ST_Centroid(geom)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

Tipo de devolución

`GEOMETRY`

Si geom es nulo, entonces se devuelve el valor nulo.

Si geom está vacío, entonces se devuelve el valor null.

Ejemplos

El siguiente SQL devuelve el punto central de una cadena de líneas de entrada de datos.

```
SELECT ST_AsEWKT(ST_Centroid(ST_GeomFromText('LINESTRING(110 40, 2 3, -10 80, -7 9, -22
 -33)', 4326)))
```

```
                     st_asewkt
-----------------------------------------------------
 SRID=4326;POINT(15.6965103455214 27.0206782881905)
```

## ST_Collect

ST_Collect tiene dos variantes. Una acepta dos geometrías y otra acepta una expresión agrupada.

La primera variante de ST_Collect crea una geometría a partir de las geometrías de entrada. El orden de las geometrías de entrada se conserva. Esta variante funciona de la siguiente manera:

- Si ambas geometrías de entrada son puntos, se devuelve un `MULTIPOINT` con dos puntos.
- Si ambas geometrías de entrada son cadenas de línea, se devuelve una `MULTILINESTRING` con dos cadenas de línea.
- Si ambas geometrías de entrada son polígonos, se devuelve un `MULTIPOLYGON` con dos polígonos.
- De lo contrario, se devuelve una `GEOMETRYCOLLECTION` con dos geometrías de entrada.

La segunda variante de ST_Collect crea una geometría a partir de las geometrías en una columna de geometría. No hay un orden de devolución de las geometrías determinado. Especifique la cláusula WITHIN GROUP (ORDER BY ...) para especificar el orden de las geometrías devueltas. Esta variante funciona de la siguiente manera:

- Si todas las filas que no son NULL de la expresión agrupada de entrada son puntos, se devuelve un multipunto que contiene todos los puntos de la expresión agrupada.
- Si todas las filas que no son NULL de la expresión agrupada son cadenas de línea, se devuelve una cadena de varias líneas que contiene todas las cadenas de línea de la expresión agrupada.
- Si todas las filas que no son NULL de la expresión agrupada son polígonos, el resultado que se devuelve es un multipolígono que contiene todos los polígonos de la expresión agrupada.
- De lo contrario, se devuelve una `GEOMETRYCOLLECTION` que contiene todas las geometrías de la expresión agrupada.

ST_Collect devuelve una geometría de la misma dimensión que las geometrías de entrada. Todas las geometrías de entrada deben ser de la misma dimensión.

Sintaxis

```
ST_Collect(geom1, geom2)
```

```
ST_Collect(aggregate_expression)  [WITHIN GROUP (ORDER BY sort_expression1 [ASC | DESC]
 [, sort_expression2 [ASC | DESC] ...])]
```

Argumentos

geom1

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

geom2

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

expresión_de_agregación

Una columna de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

[WITHIN GROUP (ORDER BY sort_expression1 [ASC | DESC] [, sort_expression2 [ASC | DESC] ...])]

Una cláusula opcional que especifica el orden de los valores agregados. La cláusula ORDER BY contiene una lista de expresiones de ordenación. Las expresiones de ordenación son expresiones

similares a las expresiones de ordenación válidas de una lista de selección de la consulta, como un nombre de columna. Puede especificar un orden ascendente (`ASC`) o descendente (`DESC`). El valor predeterminado es `ASC`.

Tipo de devolución

`GEOMETRY` de subtipo `MULTIPOINT`, `MULTILINESTRING`, `MULTIPOLYGON` o `GEOMETRYCOLLECTION`.

El valor del identificador de sistema de referencia espacial (SRID) de la geometría devuelta es el valor SRID de las geometrías de entrada de datos.

Si tanto geom1 como geom2 son null, se devuelve el valor null.

Si todas las filas de aggregate_expression son null, se devuelve el valor null.

Si geom1 es null, se devuelve una copia de geom2. Del mismo modo, si geom2 es null, se devuelve una copia de geom1.

Si geom1 y geom2 tienen diferentes valores SRID, se devuelve un error.

Si dos geometrías en aggregate_expression tienen diferentes valores SRID, se devuelve un error.

Si el tamaño de la geometría devuelta es mayor que el tamaño máximo de una `GEOMETRY`, se devuelve un error.

Si geom1 y geom2 tienen diferentes dimensiones, se devuelve un error.

Si dos geometrías en aggregate_expression tienen diferentes dimensiones, se devuelve un error.

Ejemplos

El siguiente SQL devuelve una colección de geometría que contiene las dos geometrías de entrada.

```
SELECT ST_AsText(ST_Collect(ST_GeomFromText('LINESTRING(0 0,1 1)'),
 ST_GeomFromText('POLYGON((10 10,20 10,10 20,10 10))')));
```

```
st_astext
```

```
-----------
 GEOMETRYCOLLECTION(LINESTRING(0 0,1 1),POLYGON((10 10,20 10,10 20,10 10)))
```

El siguiente SQL recopila todas las geometrías de una tabla en una colección de geometría.

```
WITH tbl(g) AS (SELECT ST_GeomFromText('POINT(1 2)', 4326) UNION ALL
SELECT ST_GeomFromText('LINESTRING(0 0,10 0)', 4326) UNION ALL
SELECT ST_GeomFromText('MULTIPOINT(13 4,8 5,4 4)', 4326) UNION ALL
SELECT NULL::geometry UNION ALL
SELECT ST_GeomFromText('POLYGON((0 0,10 0,0 10,0 0))', 4326))
SELECT ST_AsEWKT(ST_Collect(g)) FROM tbl;
```

```
st_astext
-----------
 SRID=4326;GEOMETRYCOLLECTION(POINT(1 2),LINESTRING(0 0,10 0),MULTIPOINT((13 4),(8 5),
(4 4)),POLYGON((0 0,10 0,0 10,0 0)))
```

El siguiente SQL recopila todas las geometrías de la tabla agrupadas por la columna id y ordenadas por este ID. En este ejemplo, las geometrías resultantes se agrupan por ID de la siguiente manera:

- id 1: puntos en un multipunto.
- id 2: cadenas de línea en una cadena de varias líneas.
- id 3: subtipos mixtos en una colección de geometría.
- id 4: polígonos en un multipolígono.
- id 5: null y el resultado es null.

```
WITH tbl(id, g) AS (SELECT 1, ST_GeomFromText('POINT(1 2)', 4326) UNION ALL
SELECT 1, ST_GeomFromText('POINT(4 5)', 4326) UNION ALL
SELECT 2, ST_GeomFromText('LINESTRING(0 0,10 0)', 4326) UNION ALL
SELECT 2, ST_GeomFromText('LINESTRING(10 0,20 -5)', 4326) UNION ALL
SELECT 3, ST_GeomFromText('MULTIPOINT(13 4,8 5,4 4)', 4326) UNION ALL
SELECT 3, ST_GeomFromText('MULTILINESTRING((-1 -1,-2 -2),(-3 -3,-5 -5))', 4326) UNION
 ALL
SELECT 4, ST_GeomFromText('POLYGON((0 0,10 0,0 10,0 0))', 4326) UNION ALL
SELECT 4, ST_GeomFromText('POLYGON((20 20,20 30,30 20,20 20))', 4326) UNION ALL
SELECT 1, NULL::geometry UNION ALL SELECT 2, NULL::geometry UNION ALL
SELECT 5, NULL::geometry UNION ALL SELECT 5, NULL::geometry)
```

```
SELECT id, ST_AsEWKT(ST_Collect(g)) FROM tbl GROUP BY id ORDER BY id;
```

```
 id |                                                st_asewkt

----
+-----------------------------------------------------------------------------------------------
  1 | SRID=4326;MULTIPOINT((1 2),(4 5))
  2 | SRID=4326;MULTILINESTRING((0 0,10 0),(10 0,20 -5))
  3 | SRID=4326;GEOMETRYCOLLECTION(MULTIPOINT((13 4),(8 5),(4 4)),MULTILINESTRING((-1
 -1,-2 -2),(-3 -3,-5 -5)))
  4 | SRID=4326;MULTIPOLYGON(((0 0,10 0,0 10,0 0)),((20 20,20 30,30 20,20 20)))
  5 |
```

El siguiente SQL recopila todas las geometrías de una tabla en una colección de geometría. Los resultados se ordenan de forma descendente por `id` y, luego, lexicográficamente, en función de sus coordenadas x mínimas y máximas.

```
WITH tbl(id, g) AS (
SELECT 1, ST_GeomFromText('POINT(4 5)', 4326) UNION ALL
SELECT 1, ST_GeomFromText('POINT(1 2)', 4326) UNION ALL
SELECT 2, ST_GeomFromText('LINESTRING(10 0,20 -5)', 4326) UNION ALL
SELECT 2, ST_GeomFromText('LINESTRING(0 0,10 0)', 4326) UNION ALL
SELECT 3, ST_GeomFromText('MULTIPOINT(13 4,8 5,4 4)', 4326) UNION ALL
SELECT 3, ST_GeomFromText('MULTILINESTRING((-1 -1,-2 -2),(-3 -3,-5 -5))', 4326) UNION
 ALL
SELECT 4, ST_GeomFromText('POLYGON((20 20,20 30,30 20,20 20))', 4326) UNION ALL
SELECT 4, ST_GeomFromText('POLYGON((0 0,10 0,0 10,0 0))', 4326) UNION ALL
SELECT 1, NULL::geometry UNION ALL SELECT 2, NULL::geometry UNION ALL
SELECT 5, NULL::geometry UNION ALL SELECT 5, NULL::geometry)
SELECT ST_AsEWKT(ST_Collect(g) WITHIN GROUP (ORDER BY id DESC, ST_XMin(g), ST_XMax(g)))
 FROM tbl;
```

```
                          st_asewkt

-----------------------------------------------------------------------------------------------
 SRID=4326;GEOMETRYCOLLECTION(POLYGON((0 0,10 0,0 10,0 0)),POLYGON((20 20,20 30,30
 20,20 20)),MULTILINESTRING((-1 -1,-2 -2),(-3 -3,-5 -5)),MULTIPOINT((13 4),(8 5),(4
 4)),LINESTRING(0 0,10 0),LINESTRING(10 0,20 -5),POINT(1 2),POINT(4 5)
```

## ST_Contains

ST_Contains devuelve true si la proyección 2D de la primera geometría de entrada contiene la proyección 2D de la segunda geometría de entrada. La geometría A contiene geometría B si cada punto de B es un punto en A, y sus interiores tienen intersecciones no vacías.

ST_Contains(A, B) es equivalente a ST_Within(B, A).

### Sintaxis

```
ST_Contains(geom1, geom2)
```

### Argumentos

geom1

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

geom2

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY. Este valor se compara con geom1 para determinar si está contenido dentro de geom1.

### Tipo de devolución

BOOLEAN

Si geom1 o geom2 son null, entonces se devuelve el valor null.

Si geom1 y geom2 no tienen el mismo valor en el identificador de sistema de referencia espacial (SRID), se devuelve un error.

Si geom1 o geom2 son una colección geométrica, se devuelve un error.

### Ejemplos

El siguiente SQL revisa si el primer polígono contiene el segundo polígono.

```
SELECT ST_Contains(ST_GeomFromText('POLYGON((0 2,1 1,0 -1,0 2))'),
 ST_GeomFromText('POLYGON((-1 3,2 1,0 -3,-1 3))'));
```

```
st_contains
```

```
-----------
 false
```

## ST_ContainsProperly

ST_ContainsProperly devuelve true si ambas geometrías de entrada no están vacías, y todos los puntos de la proyección 2D de la segunda geometría son puntos interiores de la proyección 2D de la primera geometría.

### Sintaxis

```
ST_ContainsProperly(geom1, geom2)
```

### Argumentos

geom1

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`. El subtipo no puede ser `GEOMETRYCOLLECTION`.

geom2

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`. El subtipo no puede ser `GEOMETRYCOLLECTION`. Este valor se compara con geom1 para determinar si todos sus puntos son puntos interiores de geom1.

### Tipo de devolución

`BOOLEAN`

Si geom1 o geom2 son null, entonces se devuelve el valor null.

Si geom1 y geom2 no tienen el mismo valor en el identificador de sistema de referencia espacial (SRID), se devuelve un error.

Si geom1 o geom2 son una colección geométrica, se devuelve un error.

### Ejemplos

El siguiente SQL devuelve los valores de ST_Contains y ST_ContainsProperly donde la cadena de línea de entrada forma una intersección con el interior y el límite del polígono de entrada (pero no con su exterior). El polígono contiene la cadena de línea, pero no lo hace de forma adecuada.

```
WITH tmp(g1, g2)
AS (SELECT ST_GeomFromText('POLYGON((0 0,10 0,10 10,0 10,0 0))'),
 ST_GeomFromText('LINESTRING(5 5,10 5,10 6,5 5)')) SELECT ST_Contains(g1, g2),
 ST_ContainsProperly(g1, g2)
FROM tmp;
```

```
 st_contains | st_containsproperly
-------------+---------------------
 t           | f
```

## ST_ConvexHull

ST_ConvexHull devuelve una geometría que representa la envolvente convexa de los puntos no vacíos que se incluyen en la geometría de entrada.

En el caso de una entrada vacía, la geometría resultante es la misma que la geometría de entrada. Para todas las entradas no vacías, la función opera en la proyección 2D de la geometría de entrada. No obstante, la dimensión de la geometría de salida depende de la dimensión de la geometría de entrada. Más específicamente, cuando la geometría de entrada es una geometría 3DM o 3D no vacía, se eliminan las coordenadas `m`. Es decir, la dimensión de la geometría devuelta es 2D o 3DZ, respectivamente. Si la entrada es una geometría 2D o 3DZ no vacía, la geometría resultante tiene la misma dimensión.

### Sintaxis

```
ST_ConvexHull(geom)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

### Tipo de devolución

`GEOMETRY`

El valor del identificador de sistema de referencia espacial (SRID) de la geometría devuelta es el valor SRID de la geometría de entrada.

Si geom es nulo, entonces se devuelve el valor nulo.

Los valores retornados son los siguientes.

| Número de puntos en la envolvente convexa | Subtipo de geometría. |
|---|---|
| 0 | Se devuelve una copia de geom. |
| 1 | Se devuelve un subtipo `POINT`. |
| 2 | Se devuelve un subtipo `LINESTRING` . Los dos puntos de la cadena de línea devuelta están ordenados lexicográficamente. |
| 3 o mayor | Se devuelve un subtipo `POLYGON` sin anillos interiores. El polígono está orientado en sentido horario, y el primer punto del anillo exterior es el punto lexicográficamente más pequeño del anillo. |

Ejemplos

El siguiente SQL devuelve la representación de texto conocida extendida (EWKT) de una cadena de línea. En este caso, la envolvente convexa devuelta es un polígono.

```
SELECT ST_AsEWKT(ST_ConvexHull(ST_GeomFromText('LINESTRING(0 0,1 0,0 1,1 1,0.5 0.5)')))
 as output;
```

```
output
-------------
POLYGON((0 0,0 1,1 1,1 0,0 0))
```

El siguiente SQL devuelve la representación EWKT de una linestring. En este caso, la envolvente convexa devuelta es una cadena de línea.

```
SELECT ST_AsEWKT(ST_ConvexHull(ST_GeomFromText('LINESTRING(0 0,1 1,0.2 0.2,0.6 0.6,0.5
 0.5)'))) as output;
```

```
output
-------------
LINESTRING(0 0,1 1)
```

El siguiente SQL devuelve la representación EWKT de un multipunto. En este caso, la envolvente convexa devuelta es un punto.

```
SELECT ST_AsEWKT(ST_ConvexHull(ST_GeomFromText('MULTIPOINT(0 0,0 0,0 0)'))) as output;
```

```
output
-------------
POINT(0 0)
```

## ST_CoveredBy

ST_CoveredBy devuelve true si la proyección 2D de la primera geometría de entrada está cubierta por la proyección 2D de la segunda geometría de entrada. La geometría A está cubierta por la geometría B si ninguna de ellas está vacía y cada punto de A es un punto en B.

ST_CoveredBy(A, B) es equivalente a ST_Covers(B, A).

### Sintaxis

```
ST_CoveredBy(geom1, geom2)
```

### Argumentos

geom1

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY. Este valor se compara con geom2 para determinar si está cubierto por geom2.

geom2

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

### Tipo de devolución

BOOLEAN

Si geom1 o geom2 son null, entonces se devuelve el valor null.

Si geom1 y geom2 no tienen el mismo valor en el identificador de sistema de referencia espacial (SRID), se devuelve un error.

Si geom1 o geom2 son una colección geométrica, se devuelve un error.

Ejemplos

El siguiente SQL revisa si el primer polígono está cubierto por el segundo polígono.

```
SELECT ST_CoveredBy(ST_GeomFromText('POLYGON((0 2,1 1,0 -1,0 2))'),
 ST_GeomFromText('POLYGON((-1 3,2 1,0 -3,-1 3))'));
```

```
st_coveredby
-----------
 true
```

## ST_Covers

ST_Covers devuelve true si la proyección 2D de la primera geometría de entrada cubre la proyección 2D de la segunda geometría de entrada. La geometría A cubre la geometría B si ninguna de ellas está vacía y cada punto de B es un punto en A.

ST_Covers(A, B) es equivalente a ST_CoveredbBy(B, A).

Sintaxis

```
ST_Covers(geom1, geom2)
```

Argumentos

geom1

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

geom2

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY. Este valor se compara con geom1 para determinar si cubre geom1.

Tipo de devolución

`BOOLEAN`

Si geom1 o geom2 son null, entonces se devuelve el valor null.

Si geom1 y geom2 no tienen el mismo valor en el identificador de sistema de referencia espacial (SRID), se devuelve un error.

Si geom1 o geom2 son una colección geométrica, se devuelve un error.

Ejemplos

El siguiente SQL revisa si el primer polígono cubre el segundo polígono.

```
SELECT ST_Covers(ST_GeomFromText('POLYGON((0 2,1 1,0 -1,0 2))'),
 ST_GeomFromText('POLYGON((-1 3,2 1,0 -3,-1 3))'));
```

```
st_covers
-----------
 false
```

## ST_Crosses

ST_Crosses devuelve true si las proyecciones 2D de las dos geometrías de entrada se cruzan entre sí.

Sintaxis

```
ST_Crosses(geom1, geom2)
```

Argumentos

geom1

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

geom2

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

Tipo de devolución

`BOOLEAN`

Si geom1 o geom2 son null, entonces se devuelve un error.

Si geom1 o geom2 son una colección geométrica, se devuelve un error.

Si geom1 y geom2 no tienen el mismo valor en el identificador de sistema de referencia espacial (SRID), se devuelve un error.

Ejemplos

El siguiente SQL verifica si el primer polígono se cruza con el segundo multipunto. En este ejemplo, el multipunto forma una intersección tanto con el interior del polígono como con su exterior, razón por la que ST_Crosses devuelve true.

```
SELECT ST_Crosses (ST_GeomFromText('polygon((0 0,10 0,10 10,0 10,0 0))'),
 ST_GeomFromText('multipoint(5 5,0 0,-1 -1)'));
```

```
st_crosses
-------------
 true
```

El siguiente SQL verifica si el primer polígono se cruza con el segundo multipunto. En este ejemplo, el multipunto forma una intersección con el exterior del polígono, pero no con su interior, razón por la que ST_Crosses devuelve false.

```
SELECT ST_Crosses (ST_GeomFromText('polygon((0 0,10 0,10 10,0 10,0 0))'),
 ST_GeomFromText('multipoint(0 0,-1 -1)'));
```

```
st_crosses
-------------
 false
```

## ST_Dimension

ST_Dimension devuelve la dimensión inherente de una geometría de datos de entrada. El inherent dimension es el valor dimensional del subtipo que se define en la geometría.

En el caso de las entradas de geometría 3DM, 3DZ y 4D, ST_Dimension devuelve el mismo resultado que para las entradas de geometría 2D.

Sintaxis

```
ST_Dimension(geom)
```

Argumentos

geom

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

Tipo de devolución

INTEGER representa la dimensión inherente de geom.

Si geom es nulo, entonces se devuelve el valor nulo.

Los valores retornados son los siguientes.

| Valor devuelto | Subtipo de geometría. |
|---|---|
| 0 | Se devuelve si geom es un subtipo de POINT o MULTIPOINT |
| 1 | Se devuelve si geom es un subtipo de LINESTRING o MULTILINESTRING . |
| 2 | Se devuelve si geom es un subtipo de POLYGON o MULTIPOLYGON |
| 0 | Se devuelve si geom es un subtipo de GEOMETRYCOLLECTION vacío |
| La dimensión más grande de los componentes de la colección | Se devuelve si geom es un subtipo de GEOMETRYCOLLECTION |

Ejemplos

El siguiente SQL convierte una representación de texto conocida (WKT) de un LINSESTRING de cuatro puntos en un objeto GEOMETRY y devuelve la dimensión del linestring.

```
SELECT ST_Dimension(ST_GeomFromText('LINESTRING(77.29 29.07,77.42 29.26,77.27
 29.31,77.29 29.07)'));
```

```
st_dimension
-------------
1
```

## ST_Disjoint

ST_Disjoint devuelve true si las proyecciones 2D de las dos geometrías de entrada no tienen ningún punto en común.

Sintaxis

```
ST_Disjoint(geom1, geom2)
```

Argumentos

geom1

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

geom2

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

Tipo de devolución

BOOLEAN

Si geom1 o geom2 son null, entonces se devuelve el valor null.

Si geom1 y geom2 no tienen el mismo valor en el identificador de sistema de referencia espacial (SRID), se devuelve un error.

Si geom1 o geom2 son una colección geométrica, se devuelve un error.

Ejemplos

El siguiente SQL revisa si el primer polígono está disociado del segundo polígono.

```
SELECT ST_Disjoint(ST_GeomFromText('POLYGON((0 0,10 0,10 10,0 10,0 0),(2 2,2 5,5 5,5
 2,2 2))'), ST_Point(4, 4));
```

```
st_disjoint
-----------
 true
```

## ST_Distance

En el caso de las geometrías de entrada, ST_Distance devuelve la distancia euclidiana mínima entre las proyecciones 2D de los dos valores de geometría de entrada.

En el caso de las geometrías 3DM, 3DZ y 4D, ST_Distance devuelve la distancia euclidiana entre las proyecciones 2D de dos valores de geometría de entrada.

En el caso de las geografías de entrada, ST_Distance devuelve la distancia geodésica de los dos puntos 2D. La unidad de distancia se expresa en metros. En el caso de las geografías distintas de los puntos y los puntos vacíos, se devuelve un error.

Sintaxis

```
ST_Distance(geo1, geo2)
```

Argumentos

geo1

Un valor de tipo de datos GEOMETRY o GEOGRAPHY, o una expresión que toma el valor de un tipo GEOMETRY o GEOGRAPHY. El tipo de datos de geo1 debe coincidir con el de geo2.

geo2

Un valor de tipo de datos GEOMETRY o GEOGRAPHY, o una expresión que toma el valor de un tipo GEOMETRY o GEOGRAPHY. El tipo de datos de geo2 debe coincidir con el de geo1.

Tipo de devolución

`DOUBLE PRECISION` en las mismas unidades que las geometrías o las geografías de entrada de datos.

Si geo1 o geo2 es nulo o está vacío, se devuelve el valor nulo.

Si geo1 y geo2 no tienen el mismo valor en el identificador de sistema de referencia espacial (SRID), se devuelve un error.

Si geo1 o geo2 es una colección geométrica, se devuelve un error.

Ejemplos

El siguiente SQL devuelve la distancia entre dos polígonos.

```
SELECT ST_Distance(ST_GeomFromText('POLYGON((0 2,1 1,0 -1,0 2))'),
 ST_GeomFromText('POLYGON((-1 -3,-2 -1,0 -3,-1 -3))'));
```

```
  st_distance
-----------
1.4142135623731
```

El siguiente SQL devuelve la distancia (en metros) entre la Puerta de Brandeburgo y el edificio del Reichstag de Berlín mediante un tipo de datos GEOGRAPHY.

```
SELECT ST_Distance(ST_GeogFromText('POINT(13.37761826722198 52.516411678282445)'),
 ST_GeogFromText('POINT(13.377950831464005 52.51705102546893)'));
```

```
   st_distance
------------------
 74.64129172609631
```

## ST_DistanceSphere

ST_DistanceSphere devuelve la distancia entre dos geometrías de puntos que están en una esfera.

### Sintaxis

```
ST_DistanceSphere(geom1, geom2)
```

```
ST_DistanceSphere(geom1, geom2, radius)
```

### Argumentos

geom1

Un valor de punto en grados de tipo de datos GEOMETRY que está en una esfera. La primera coordenada del punto es el valor de la longitud. La segunda coordinada del punto es el valor de la latitud. En el caso de las geometrías 3DZ, 3DM o 4D, solo se utilizan las dos primeras coordenadas.

geom2

Un valor de punto en grados de tipo de datos GEOMETRY que está en una esfera. La primera coordenada del punto es el valor de la longitud. La segunda coordinada del punto es el valor de la latitud. En el caso de las geometrías 3DZ, 3DM o 4D, solo se utilizan las dos primeras coordenadas.

radius

El radio de una esfera de tipo de datos DOUBLE PRECISION. Si no se proporciona radius, la esfera por defecto es la Tierra y el radio se computa a partir de la representación del sistema geodésico mundial (WGS) 84 del elipsoide.

### Tipo de devolución

DOUBLE PRECISION en las mismas unidades que el radio. Si no se proporciona ningún radio, la distancia se indica en metros.

Si geom1 o geom2 son null o empty, entonces se devuelve el valor null.

Si no se proporciona ningún radius, el resultado en metros es la superficie de la Tierra.

Si radius es un número negativo, se devuelve un error.

Si geom1 y geom2 no tienen el mismo valor en el identificador de sistema de referencia espacial (SRID), se devuelve un error.

Si geom1 o geom2 no son un punto, se devuelve un error.

Ejemplos

En el siguiente ejemplo de SQL se calcula la distancia en kilómetros entre dos puntos de la Tierra.

```
SELECT ROUND(ST_DistanceSphere(ST_Point(-122, 47), ST_Point(-122.1, 47.1))/ 1000, 0);
```

```
  round
-----------
 13
```

El siguiente SQL de ejemplo computa la distancia en kilómetros entre las ubicaciones de tres aeropuertos en Alemania: Berlin Tegel (TXL), Munich International (MUC) y Frankfurt International (FRA).

```
WITH airports_raw(code,lon,lat) AS (
(SELECT 'MUC', 11.786111, 48.353889) UNION
(SELECT 'FRA', 8.570556, 50.033333) UNION
(SELECT 'TXL', 13.287778, 52.559722)),
airports1(code,location) AS (SELECT code, ST_Point(lon, lat) FROM airports_raw),
airports2(code,location) AS (SELECT * from airports1)
SELECT (airports1.code || ' <-> ' || airports2.code) AS airports,
round(ST_DistanceSphere(airports1.location, airports2.location) / 1000, 0) AS
 distance_in_km
FROM airports1, airports2 WHERE airports1.code < airports2.code ORDER BY 1;
```

```
  airports   | distance_in_km
-------------+----------------
 FRA <-> MUC |            299
 FRA <-> TXL |            432
 MUC <-> TXL |            480
```

## ST_DWithin

ST_DWithin devuelve true si la distancia euclidiana entre las proyecciones 2D de los dos valores de geometría de entrada no es mayor que el valor de un límite.

### Sintaxis

```
ST_DWithin(geom1, geom2, threshold)
```

### Argumentos

geom1

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

geom2

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

threshold

Un valor de tipo de datos DOUBLE PRECISION. Este valor está en las mismas unidades que los argumentos de entrada.

### Tipo de devolución

BOOLEAN

Si geom1 o geom2 son null, entonces se devuelve el valor null.

Si threshold es negativo, se devuelve un error.

Si geom1 y geom2 no tienen el mismo valor en el identificador de sistema de referencia espacial (SRID), se devuelve un error.

Si geom1 o geom2 son una colección geométrica, se devuelve un error.

### Ejemplos

El siguiente SQL comprueba si la distancia entre dos polígonos está en un rango de cinco unidades.

```
SELECT ST_DWithin(ST_GeomFromText('POLYGON((0 2,1 1,0 -1,0 2))'),
 ST_GeomFromText('POLYGON((-1 3,2 1,0 -3,-1 3))'),5);
```

```
st_dwithin
-----------
```

```
 true
```

## ST_EndPoint

ST_Endpoint devuelve el último punto de una linestring de entrada. El valor del identificador del sistema de referencia espacial (SRID) del resultado es el mismo que el de la geometría de entrada. La dimensión de la geometría devuelta es la misma que la de la geometría de entrada.

### Sintaxis

```
ST_EndPoint(geom)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`. El subtipo debe ser `LINESTRING` .

### Tipo de devolución

`GEOMETRY`

Si geom es nulo, entonces se devuelve el valor nulo.

Si geom está vacío, entonces se devuelve el valor null.

Si geom no es `LINESTRING`, entonces se devuelve el valor null.

### Ejemplos

El siguiente SQL devuelve una representación extended well-known text (EWKT) de una `LINESTRING` de cuatro puntos a un objeto `GEOMETRY` y devuelve el punto final de la linestring.

```
SELECT ST_AsEWKT(ST_EndPoint(ST_GeomFromText('LINESTRING(0 0,10 0,10 10,5 5,0
 5)',4326)));
```

```

st_asewkt
-------------
 SRID=4326;POINT(0 5)
```

## ST_Envelope

ST_Envelope devuelve el cuadro delimitador mínimo de la geometría de entrada, de la siguiente manera:

- Si la geometría de entrada está vacía, la geometría devuelta es una copia de la geometría de entrada.
- Si el cuadro delimitador mínimo de la geometría de entrada degenera en un punto, la geometría devuelta es un punto.
- Si el cuadro delimitador mínimo de la geometría de entrada es unidimensional, se devuelve una linestring de dos puntos.
- Si ninguna de las anteriores es verdadera, la función devuelve un polígono orientado en sentido horario cuyos vértices son las esquinas del cuadro delimitador mínimo.

El identificador del sistema de referencia espacial (SRID) de la geometría devuelta es el mismo que el de la geometría de entrada.

Para todas las entradas no vacías, la función opera en la proyección 2D de la geometría de entrada.

Sintaxis

```
ST_Envelope(geom)
```

Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

Tipo de devolución

`GEOMETRY`

Si geom es nulo, entonces se devuelve el valor nulo.

Ejemplos

El siguiente SQL convierte una representación well-known text (WKT) de una `LINESTRING` de cuatro puntos en un objeto `GEOMETRY` y devuelve un polígono cuyos vértices son el cuadro delimitador mínimo.

```
SELECT ST_AsText(ST_Envelope(ST_GeomFromText('GEOMETRYCOLLECTION(POLYGON((0 0,10 0,0
 10,0 0)),LINESTRING(20 10,20 0,10 0))')));
```

```
    st_astext
-----------------------------------
  POLYGON((0 0,0 10,20 10,20 0,0 0))
```

## ST_Equals

ST_Equals devuelve true si las proyecciones 2D de las geometrías de entrada son geométricamente iguales. Las geometrías se consideran geométricamente iguales si tienen conjuntos de puntos idénticos y sus interiores tienen una intersección que no está vacía.

### Sintaxis

```
ST_Equals(geom1, geom2)
```

### Argumentos

geom1

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

geom2

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`. Este valor se compara con geom1 para determinar si es igual a geom1.

### Tipo de devolución

`BOOLEAN`

Si geom1 o geom2 son null, entonces se devuelve un error.

Si geom1 y geom2 no tienen el mismo valor en el identificador de sistema de referencia espacial (SRID), se devuelve un error.

Si geom1 o geom2 son una colección geométrica, se devuelve un error.

### Ejemplos

El siguiente SQL comprueba si los dos polígonos son geométricamente iguales.

```
SELECT ST_Equals(ST_GeomFromText('POLYGON((0 2,1 1,0 -1,0 2))'),
 ST_GeomFromText('POLYGON((-1 3,2 1,0 -3,-1 3))'));
```

```
st_equals
-----------
 false
```

El siguiente SQL comprueba si los dos linestrings son geométricamente iguales.

```
SELECT ST_Equals(ST_GeomFromText('LINESTRING(1 0,10 0)'), ST_GeomFromText('LINESTRING(1
 0,5 0,10 0)'));
```

```
st_equals
-----------
 true
```

## ST_ExteriorRing

ST_ExteriorRing devuelve una cadena de línea cerrada que representa el anillo exterior de un polígono de entrada. La dimensión de la geometría devuelta es la misma que la de la geometría de entrada.

### Sintaxis

```
ST_ExteriorRing(geom)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

Tipo de devolución

GEOMETRY del subtipo LINESTRING.

El valor del identificador de sistema de referencia espacial (SRID) de la geometría devuelta es el valor SRID de la geometría de entrada.

Si geom es nulo, entonces se devuelve el valor nulo.

Si geom no es un polígono, se devuelve null.

Si geom está vacío, se devuelve un polígono vacío.

Ejemplos

El siguiente SQL devuelve el anillo exterior de un polígono como una cadena de línea cerrada.

```
SELECT ST_AsEWKT(ST_ExteriorRing(ST_GeomFromText('POLYGON((7 9,8 7,11 6,15 8,16 6,17
 7,17 10,18 12,17 14,15 15,11 15,10 13,9 12,7 9),(9 9,10 10,11 11,11 10,10 8,9 9),(12
 14,15 14,13 11,12 14))')));
```

```
st_asewkt
-----------
 LINESTRING(7 9,8 7,11 6,15 8,16 6,17 7,17 10,18 12,17 14,15 15,11 15,10 13,9 12,7 9)
```

## ST_Force2D

ST_Force2D devuelve una geometría 2D de la geometría de entrada. En el caso de las geometrías 2D, se devuelve una copia de la entrada. En el caso de las geometrías 3DZ, 3DM y 4D, ST_Force2D proyecta la geometría en el plano cartesiano XY. Los puntos vacíos en la geometría de entrada continúan siendo puntos vacíos en la geometría de salida.

Sintaxis

```
ST_Force2D(geom)
```

Argumentos

geom

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

Tipo de devolución

`GEOMETRY`.

El valor del identificador de sistema de referencia espacial (SRID) de la geometría devuelta es el valor SRID de la geometría de entrada.

Si geom es nulo, entonces se devuelve el valor nulo.

Si geom está vacío, se devuelve una geometría vacía.

Ejemplos

El siguiente SQL devuelve una geometría 2D de una geometría 3DZ.

```
SELECT ST_AsEWKT(ST_Force2D(ST_GeomFromText('MULTIPOINT Z(0 1 2, EMPTY, 2 3 4, 5 6
 7)')));
```

```
st_asewkt
-----------
  MULTIPOINT((0 1),EMPTY,(2 3),(5 6))
```

## ST_Force3D

ST_Force3D es un alias para ST_Force3DZ. Para obtener más información, consulte ST_Force3DZ.

## ST_Force3DM

ST_Force3DM devuelve una geometría 3DM de la geometría de entrada. En el caso de las geometrías 2D, las coordenadas `m` de los puntos no vacíos en la geometría de salida se establecen en `0`. En el caso de las geometrías 3DM, se devuelve una copia de la geometría de entrada.
En el caso de las geometrías 3DZ, la geometría se proyecta en el plano cartesiano XY, y las coordenadas `m` de los puntos no vacíos en la geometría de salida se establecen en `0`. En el caso de las geometrías 4D, la geometría se proyecta en el espacio cartesiano XYM. Los puntos vacíos en la geometría de entrada continúan siendo puntos vacíos en la geometría de salida.

Sintaxis

```
ST_Force3DM(geom)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

### Tipo de devolución

`GEOMETRY`.

El valor del identificador de sistema de referencia espacial (SRID) de la geometría devuelta es el valor SRID de la geometría de entrada.

Si geom es nulo, entonces se devuelve el valor nulo.

Si geom está vacío, se devuelve una geometría vacía.

### Ejemplos

El siguiente SQL devuelve una geometría 3DM de una geometría 3DZ.

```
SELECT ST_AsEWKT(ST_Force3DM(ST_GeomFromText('MULTIPOINT Z(0 1 2, EMPTY, 2 3 4, 5 6
 7)')));
```

```
st_asewkt
-----------
  MULTIPOINT M ((0 1 0),EMPTY,(2 3 0),(5 6 0))
```

## ST_Force3DZ

ST_Force3DZ devuelve una geometría 3DZ de la geometría de entrada. En el caso de las geometrías 2D, las coordenadas z de los puntos no vacíos en la geometría de salida se establecen en `0`. En el caso de las geometrías 3DM, la geometría se proyecta en el plano cartesiano XY, y las coordenadas z de los puntos no vacíos en la geometría de salida se establecen en `0`. En el caso de las geometrías 3DZ, se devuelve una copia de la geometría de entrada. En el caso de las geometrías 4D, la geometría se proyecta en el espacio cartesiano XYZ. Los puntos vacíos en la geometría de entrada continúan siendo puntos vacíos en la geometría de salida.

### Sintaxis

```
ST_Force3DZ(geom)
```

### Argumentos

geom

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

### Tipo de devolución

GEOMETRY.

El valor del identificador de sistema de referencia espacial (SRID) de la geometría devuelta es el valor SRID de la geometría de entrada.

Si geom es nulo, entonces se devuelve el valor nulo.

Si geom está vacío, se devuelve una geometría vacía.

### Ejemplos

El siguiente SQL devuelve una geometría 3DZ de una geometría 3DM.

```
SELECT ST_AsEWKT(ST_Force3DZ(ST_GeomFromText('MULTIPOINT M(0 1 2, EMPTY, 2 3 4, 5 6
 7)')));
```

```
st_asewkt
-----------
  MULTIPOINT Z ((0 1 0),EMPTY,(2 3 0),(5 6 0))
```

## ST_Force4D

ST_Force4D devuelve una geometría 4D de la geometría de entrada. En el caso de las geometrías 2D, las coordenadas z y m de los puntos no vacíos en la geometría de salida se establecen en 0. En el caso de las geometrías 3DM, las coordenadas z de los puntos no vacíos en la geometría de salida se establecen en 0. En el caso de las geometrías 3DZ, las coordenadas m de los puntos no vacíos en la geometría de salida se establecen en 0. En el caso de las geometrías 4D, se devuelve una copia

de la geometría de entrada. Los puntos vacíos en la geometría de entrada continúan siendo puntos vacíos en la geometría de salida.

Sintaxis

```
ST_Force4D(geom)
```

Argumentos

geom

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

Tipo de devolución

GEOMETRY.

El valor del identificador de sistema de referencia espacial (SRID) de la geometría devuelta es el valor SRID de la geometría de entrada.

Si geom es nulo, entonces se devuelve el valor nulo.

Si geom está vacío, se devuelve una geometría vacía.

Ejemplos

El siguiente SQL devuelve una geometría 4D a partir de una geometría 3DM.

```
SELECT ST_AsEWKT(ST_Force4D(ST_GeomFromText('MULTIPOINT M(0 1 2, EMPTY, 2 3 4, 5 6
 7)')));
```

```
st_asewkt
-----------
  MULTIPOINT ZM ((0 1 0 2),EMPTY,(2 3 0 4),(5 6 0 7))
```

## ST_GeoHash

ST_GeoHash devuelve la representación geohash del punto de entrada con la precisión especificada. El valor de precisión predeterminado es de 20. Para obtener más información sobre la definición de geohash, consulte Geohash en Wikipedia.

### Sintaxis

```
ST_GeoHash(geom)
```

```
ST_GeoHash(geom, precision)
```

### Argumentos

geom

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

precisión

Un valor de tipo de datos INTEGER. El valor predeterminado es 20.

### Tipo de devolución

GEOMETRY

La función devuelve la representación geohash del punto de entrada.

Si el punto de entrada está vacío, la función devuelve un valor nulo.

Si la geometría de entrada no es un punto, la función devuelve un error.

### Ejemplos

El siguiente SQL devuelve la representación geohash del punto de entrada.

```
SELECT ST_GeoHash(ST_GeomFromText('POINT(45 -45)'), 25) AS geohash;
```

```
         geohash
---------------------------
 m0000000000000000000gzz
```

El siguiente SQL devuelve un valor nulo porque el punto de entrada está vacío.

```
SELECT ST_GeoHash(ST_GeomFromText('POINT EMPTY'), 10) IS NULL AS result;
```

```
 result
```

```
---------
 true
```

## ST_GeogFromText

ST_GeogFromText construye un objeto geográfico a partir de una representación de texto conocido (WKT) o de texto conocido extendido (EWKT) de una geografía de entrada.

### Sintaxis

```
ST_GeogFromText(wkt_string)
```

### Argumentos

wkt_string

Un valor de tipo de datos VARCHAR que es una representación WKT o EWKT de una geografía.

### Tipo de devolución

GEOGRAPHY

Si el valor SRID se establece en el valor proporcionado en la entrada. Si no se proporciona el SRID, se establece como 4326.

Si wkt_string es nulo, se devuelve el valor nulo.

Si wkt_string no es válido, se devuelve un error.

### Ejemplos

El siguiente SQL construye un polígono a partir de un objeto geográfico cuyo valor es el SRID.

```
SELECT ST_AsEWKT(ST_GeogFromText('SRID=4324;POLYGON((0 0,0 1,1 1,10 10,1 0,0 0))'));
```

```
  st_asewkt
----------------------------------------------
 SRID=4324;POLYGON((0 0,0 1,1 1,10 10,1 0,0 0))
```

El siguiente SQL construye un polígono a partir de un objeto geográfico. El valor SRID se establece en 4326.

```
SELECT ST_AsEWKT(ST_GeogFromText('POLYGON((0 0,0 1,1 1,10 10,1 0,0 0))'));
```

```
 st_asewkt
-----------------------------------------------
 SRID=4326;POLYGON((0 0,0 1,1 1,10 10,1 0,0 0))
```

## ST_GeogFromWKB

ST_GeogFromWKB construye un objeto geográfico a partir de la representación binaria conocida (WKB) hexadecimal de una geografía de datos de entrada.

Sintaxis

```
ST_GeogFromWKB(wkb_string)
```

Argumentos

wkb_string

Un valor de tipo de datos VARCHAR que es una representación WKB hexadecimal de una geografía.

Tipo de devolución

GEOGRAPHY

Si se proporciona el valor SRID, se establece en el mismo valor proporcionado. Si no se proporciona el SRID, se establece como 4326.

Si wkb_string es nulo, entonces se devuelve el valor nulo.

Si wkb_string no es válido, se devuelve un error.

Ejemplos

El siguiente SQL construye una geografía a partir de un valor WKB hexadecimal.

```
SELECT
 ST_AsEWKT(ST_GeogFromWKB('0103000000010000000500000000000000000000000000000000000000000000000000000
```

```
 st_asewkt
-----------------------------------------
 SRID=4326;POLYGON((0 0,0 1,1 1,1 0,0 0))
```

## ST_GeometryN

ST_Geometryn devuelve una geometría a la que apunta el índice de entrada de la geometría de entrada, de la siguiente manera:

- Si la entrada es un punto, una linestring o un polígono, se devuelve una geometría tal cual si el índice es igual a uno (1) y null si el índice es distinto de uno (1).
- Si la entrada es un multipunto, multilinestring, multipolígono o colección de geometría, se devuelve un punto, linestring, polígono o colección de geometría a la que apunta un índice de entrada.

El índice tiene base uno. El identificador del sistema de referencia espacial (SRID) del resultado es el mismo que el de la geometría de entrada. La dimensión de la geometría devuelta es la misma que la de la geometría de entrada.

### Sintaxis

```
ST_GeometryN(geom, index)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

índice

Un valor de tipo de datos `INTEGER` que representa la posición de un índice con base 1.

### Tipo de devolución

`GEOMETRY`

Si geom o index es null, entonces se devuelve el valor null.

Si el índice está fuera de rango, se devuelve un error.

Ejemplos

El siguiente SQL devuelve las geometrías en una colección de geometría.

```
WITH tmp1(idx) AS (SELECT 1 UNION SELECT 2),
tmp2(g) AS (SELECT ST_GeomFromText('GEOMETRYCOLLECTION(POLYGON((0 0,10 0,0 10,0
 0)),LINESTRING(20 10,20 0,10 0))'))
SELECT idx, ST_AsEWKT(ST_GeometryN(g, idx)) FROM tmp1, tmp2 ORDER BY idx;
```

```
 idx |          st_asewkt
-----+-----------------------------
   1 | POLYGON((0 0,10 0,0 10,0 0))
   2 | LINESTRING(20 10,20 0,10 0)
```

## ST_GeometryType

ST_GeometryType devuelve el subtipo de una geometría de entrada de datos como una cadena.

En el caso de las entradas de las geometrías 3DM, 3DZ y 4D, ST_GeometryType devuelve el mismo resultado que para las entradas de la geometría 2D.

Sintaxis

```
ST_GeometryType(geom)
```

Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

Tipo de devolución

`VARCHAR` representa el subtipo de geom.

Si geom es nulo, entonces se devuelve el valor nulo.

Los valores retornados son los siguientes.

| Valor de cadena devuelto | Subtipo de geometría. |
|---|---|
| ST_Point | Se devuelve si geom es un subtipo de POINT |
| ST_LineString | Se devuelve si geom es un subtipo de LINESTRING |
| ST_Polygon | Se devuelve si geom es un subtipo de POLYGON |
| ST_MultiPoint | Se devuelve si geom es un subtipo de MULTIPOINT |
| ST_MultiLineString | Se devuelve si geom es un subtipo de MULTILINESTRING |
| ST_MultiPolygon | Se devuelve si geom es un subtipo de MULTIPOLYGON |
| ST_GeometryCollection | Se devuelve si geom es un subtipo de GEOMETRYCOLLECTION |

Ejemplos

El siguiente SQL devuelve el subtipo de la geometría de linestring de entrada de datos.

```
SELECT ST_GeometryType(ST_GeomFromText('LINESTRING(77.29 29.07,77.42 29.26,77.27
 29.31,77.29 29.07)'));
```

```
st_geometrytype
-------------
 ST_LineString
```

# ST_GeomFromEWKB

ST_GeomFromEWKB construye un objeto geométrico a partir de la representación binaria conocida extendida (EWKB) de una geometría de datos de entrada.

ST_GeomFromEWKB acepta geometrías 3DZ, 3DM y 4D escritas en formato hexadecimal WKB y EWKB.

## Sintaxis

```
ST_GeomFromEWKB(ewkb_string)
```

## Argumentos

ewkb_string

Un valor de tipo de datos VARCHAR que es una representación hexadecimal EWKB de una geometría.

## Tipo de devolución

GEOMETRY

Si ewkb_string es null, entonces se devuelve el valor null.

Si ewkb_string no es válido, se devuelve un error.

## Ejemplos

El siguiente SQL construye un polígono a partir de un valor EWKB y devuelve la representación EWKT de un polígono.

```
SELECT
 ST_AsEWKT(ST_GeomFromEWKB('0103000020E6100000010000000500000000000000000000000000000000000000000000
```

```
 st_asewkt
-------------------------------
 SRID=4326;POLYGON((0 0,0 1,1 1,1 0,0 0))
```

# ST_GeomFromEWKT

ST_GeomFromEWKT construye un objeto geométrico a partir de la representación de texto conocida extendida (EWKT) de una geometría de entrada.

ST_GeomFromEWKT acepta 3DZ, 3DM y 4D donde el tipo de geometría tiene el prefijo Z, M o ZM, respectivamente.

## Sintaxis

```
ST_GeomFromEWKT(ewkt_string)
```

## Argumentos

ewkt_string

Un valor del tipo de datos `VARCHAR` o una expresión que toma el valor de un tipo `VARCHAR`, es decir, una representación EWKT de una geometría.

Puede usar la palabra clave de WKT `EMPTY` para designar un punto vacío, un multipunto con un punto vacío o una colección de geometría con un punto vacío. En el siguiente ejemplo, se crea un punto vacío.

```
ST_GeomFromEWKT('SRID=4326;POINT EMPTY');
```

## Tipo de devolución

`GEOMETRY`

Si ewkt_string es null, se devuelve el valor null.

Si ewkt_string no es válido, se devuelve un error.

## Ejemplos

El siguiente SQL crea una cadena de varias líneas a partir de un valor EWKT y devuelve una geometría. También devuelve el resultado ST_AsEWKT de la geometría.

```
SELECT ST_GeomFromEWKT('SRID=4326;MULTILINESTRING((1 0,1 0),(2 0,3 0),(4 0,5 0,6 0))')
 as geom, ST_AsEWKT(geom);
```

```
                                                    geom

                                          |                         st_asewkt

-----------------------------------------------------------------------------------------------
+-----------------------------------------------------------
 0105000020E6100000030000000102000000020000000000000000F03F0000000000000000000000000000F03F00
 | SRID=4326;MULTILINESTRING((1 0,1 0),(2 0,3 0),(4 0,5 0,6 0))
```

## ST_GeomFromGeoHash

ST_GeomFromGeoHash construye un objeto geométrico a partir de la representación geohash de una geometría de entrada. ST_GeomFromGeoHash devuelve una geometría bidimensional (2D) con el identificador de referencia espacial (SRID) cero (0). Para obtener más información sobre el formato geohash, consulte Geohash en la Wikipedia.

### Sintaxis

```
ST_GeomFromGeoHash(geohash_string)
```

```
ST_GeomFromGeoHash(geohash_string, precision)
```

### Argumentos

geohash_string

Un valor del tipo de datos VARCHAR o una expresión que toma el valor de un tipo VARCHAR, es decir, una representación geohash de una geometría.

precisión

Un valor de tipo de datos INTEGER que representa la precisión del geohash. El valor es el número de caracteres del geohash que se utilizará como precisión. Si no se especifica el valor, es inferior a cero o es superior a la longitud de geohash_string, se utiliza la longitud de geohash_string.

### Tipo de devolución

GEOMETRY

Si geohash_string es nulo, se devuelve el valor nulo.

Si geohash_string no es válido, se devuelve un error.

Ejemplos

El siguiente código SQL devuelve un polígono con precisión alta.

```
SELECT ST_AsText(ST_GeomFromGeoHash('9qqj7nmxncgyy4d0dbxqz0'));
```

```
 st_asewkt
-----------------------
 POLYGON((-115.172816 36.114646,-115.172816 36.114646,-115.172816 36.114646,-115.172816
 36.114646,-115.172816 36.114646))
```

El siguiente código SQL devuelve un punto con precisión alta.

```
SELECT ST_AsText(ST_GeomFromGeoHash('9qqj7nmxncgyy4d0dbxqz00'));
```

```
 st_asewkt
-----------------------
 POINT(-115.172816 36.114646)
```

El siguiente código SQL devuelve un polígono con precisión baja.

```
SELECT ST_AsText(ST_GeomFromGeoHash('9qq'));
```

```
 st_asewkt
-----------------------
 POLYGON((-115.3125 35.15625,-115.3125 36.5625,-113.90625 36.5625,-113.90625
 35.15625,-115.3125 35.15625))
```

El siguiente código SQL devuelve un polígono con una precisión de 3.

```
SELECT ST_AsText(ST_GeomFromGeoHash('9qqj7nmxncgyy4d0dbxqz0', 3));
```

```
 st_asewkt
-----------------------
 POLYGON((-115.3125 35.15625,-115.3125 36.5625,-113.90625 36.5625,-113.90625
 35.15625,-115.3125 35.15625))
```

## ST_GeomFromGeoJSON

ST_GeomFromGeoJSON construye un objeto geométrico a partir de la representación GeoJSON de una geometría de entrada. Para obtener más información sobre el formato GeoJSON, consulte GeoJSON en la Wikipedia.

Si hay al menos un punto con tres o más coordenadas, la geometría resultante es 3DZ, donde el componente Z es cero para los puntos que solo tienen dos coordenadas. Si todos los puntos del GeoJSON de entrada contienen dos coordenadas o están vacíos, ST_GeomFromGeojSON devuelve una geometría 2D. La geometría devuelta siempre tiene el identificador de referencia espacial (SRID) 4326.

Sintaxis

```
ST_GeomFromGeoJSON(geojson_string)
```

Argumentos

geojson_string

Un valor del tipo de datos VARCHAR o una expresión que toma el valor de un tipo VARCHAR, es decir, una representación GeoJSON de una geometría.

Tipo de devolución

GEOMETRY

Si geojson_string es nulo, se devuelve el valor nulo.

Si geojson_string no es válido, se devuelve un error.

### Ejemplos

El siguiente SQL devuelve una geometría 2D representada en el GeoJSON de entrada.

```
SELECT ST_AsEWKT(ST_GeomFromGeoJSON('{"type":"Point","coordinates":[1,2]}'));
```

```
 st_asewkt
-----------------------
 SRID=4326;POINT(1 2)
```

El siguiente SQL devuelve una geometría 3DZ representada en el GeoJSON de entrada.

```
SELECT ST_AsEWKT(ST_GeomFromGeoJSON('{"type":"LineString","coordinates":[[1,2,3],
[4,5,6],[7,8,9]]}'));
```

```
 st_asewkt
-----------------------------------------
 SRID=4326;LINESTRING Z (1 2 3,4 5 6,7 8 9)
```

El siguiente SQL devuelve la geometría 3DZ cuando solo un punto tiene tres coordenadas mientras los demás tienen dos en el GeoJSON de entrada.

```
SELECT ST_AsEWKT(ST_GeomFromGeoJSON('{"type":"Polygon","coordinates":[[[0, 0],[0, 1,
 8],[1, 0],[0, 0]]]}'));
```

```
 st_asewkt
-------------------------------------------------
 SRID=4326;POLYGON Z ((0 0 0,0 1 8,1 0 0,0 0 0))
```

## ST_GeomFromGeoSquare

ST_GeomFromGeoSquare devuelve una geometría que cubre el área representada por un valor geosquare de entrada. La geometría devuelta siempre es bidimensional. Para calcular un valor geosquare, consulte ST_GeoSquare.

### Sintaxis

```
ST_GeomFromGeoSquare(geosquare)
```

```
ST_GeomFromGeoSquare(geosquare, max_depth)
```

### Argumentos

geosquare

Valor de tipo de datos BIGINT o expresión que se evalúa como un tipo BIGINT que es un valor geosquare que describe la secuencia de subdivisiones realizadas sobre el dominio inicial para alcanzar el cuadrado deseado. Este valor se calcula mediante ST_GeoSquare.

max_depth

Valor de tipo de datos INTEGER que representa el número máximo de subdivisiones de dominio realizadas en el dominio inicial. El valor debe ser mayor o igual a 1.

### Tipo de devolución

GEOMETRY

Si geosquare no es válido, la función devuelve un error.

Si la entrada max_depth no está dentro del intervalo, la función devuelve un error.

### Ejemplos

El siguiente código SQL devuelve una geometría a partir de un valor geosquare.

```
SELECT ST_AsText(ST_GeomFromGeoSquare(797852));
```

```
 st_astext
-----------------------------------------------------------------------------------------------
 POLYGON((13.359375 52.3828125,13.359375 52.734375,13.7109375 52.734375,13.7109375
 52.3828125,13.359375 52.3828125))
```

El siguiente código SQL devuelve una geometría a partir de un valor geosquare y una profundidad máxima de 3.

```
SELECT ST_AsText(ST_GeomFromGeoSquare(797852, 3));
```

```
 st_astext
-------------------------------------
 POLYGON((0 45,0 90,45 90,45 45,0 45))
```

El siguiente código SQL calcula primero el valor geosquare de Seattle mediante la especificación de la coordenada “x” como longitud y la coordenada “y” como latitud (-122,3, 47,6). A continuación, devuelve el polígono del geosquare. Aunque la salida es una geometría bidimensional, puede utilizarse para calcular datos espaciales en términos de longitud y latitud.

```
SELECT ST_AsText(ST_GeomFromGeoSquare(ST_GeoSquare(ST_Point(-122.3, 47.6))));
```

```
 st_astext
-------------------------------------------------------------------------------------------------
POLYGON((-122.335167014971 47.6080129947513,-122.335167014971
 47.6080130785704,-122.335166931152 47.6080130785704,-122.335166931152
 47.6080129947513,-122.335167014971 47.6080129947513))
```

## ST_GeomFromText

ST_GeomFromText construye un objeto geométrico a partir de la representación de texto conocida (WKT) de una geometría de datos de entrada.

ST_GeomFromText acepta 3DZ, 3DM y 4D donde el tipo de geometría tiene el prefijo Z, M o ZM, respectivamente.

### Sintaxis

```
ST_GeomFromText(wkt_string)
```

```
ST_GeomFromText(wkt_string, srid)
```

### Argumentos

wkt_string

Un valor de tipo de datos VARCHAR que es una representación WKT de una geometría.

Puede usar la palabra clave de WKT EMPTY para designar un punto vacío, un multipunto con un punto vacío o una colección de geometría con un punto vacío. En el siguiente ejemplo, se crea un multipunto con un punto vacío y un punto no vacío.

```
ST_GeomFromEWKT('MULTIPOINT(1 0,EMPTY)');
```

srid

Un valor de tipo de datos INTEGER que es un identificador de referencia espacial (SRID). Si se proporciona un valor SRID, la geometría devuelta tiene este valor SRID. De lo contrario, el valor SRID de la geometría devuelta se establece en cero (0).

### Tipo de devolución

GEOMETRY

Si wkt_string o srid son null, entonces se devuelve el valor null.

Si srid es negativo, se devuelve null.

Si wkt_string no es válido, se devuelve un error.

Si srid no es válido, se devuelve un error.

### Ejemplos

El siguiente SQL construye un objeto geométrico a partir de la representación WKT y el valor SRID.

```
SELECT ST_GeomFromText('POLYGON((0 0,0 1,1 1,1 0,0 0))',4326);
```

```
st_geomfromtext
--------------------------------
```

```
0103000020E6100000010000000500000000000000000000000000000000000000000000000000000000000000000000000000000000F
```

## ST_GeomFromWKB

ST_GeomFromWKB construye un objeto geométrico a partir de la representación binaria conocida (WKB) hexadecimal de una geometría de datos de entrada.

ST_GeomFromWKB acepta geometrías 3DZ, 3DM y 4D escritas en formato hexadecimal WKB.

### Sintaxis

```
ST_GeomFromWKB(wkb_string)
```

```
ST_GeomFromWKB(wkb_string, srid)
```

### Argumentos

wkb_string

Un valor de tipo de datos `VARCHAR` que es una representación hexadecimal WKB de una geometría.

srid

Un valor de tipo de datos `INTEGER` que es un identificador de referencia espacial (SRID). Si se proporciona un valor SRID, la geometría devuelta tiene este valor SRID. De lo contrario, el valor SRID de la geometría devuelta se establece en 0.

### Tipo de devolución

`GEOMETRY`

Si wkb_string o srid son null, entonces se devuelve el valor null.

Si srid es negativo, se devuelve null.

Si wkb_string no es válido, se devuelve un error.

Si srid no es válido, se devuelve un error.

### Ejemplos

El siguiente SQL construye un polígono a partir de un valor WKB y devuelve la representación WKT de un polígono.

```
SELECT
 ST_AsText(ST_GeomFromWKB('010300000001000000050000000000000000000000000000000000000000000000000000000000000000
```

```
 st_astext
-------------------------------
 POLYGON((0 0,0 1,1 1,1 0,0 0))
```

## ST_GeoSquare

ST_GeoSquare subdivide recursivamente el dominio ([-180, 180], [-90, 90]) en regiones cuadradas iguales llamadas geosquares hasta una profundidad especificada. La subdivisión se basa en la ubicación de un punto proporcionado. Uno de los geosquares que contiene el punto se subdivide en cada paso hasta alcanzar la profundidad máxima. La selección de este geosquare es estable, es decir, el resultado de la función depende solo de los argumentos de entrada. La función devuelve un valor único que identifica el geosquare final en el que se encuentra el punto.

ST_GeoSquare acepta un PUNTO en el que la coordenada “x” representa la longitud y la coordenada “y” representa la latitud. La longitud y la latitud están limitadas a [-180, 180] y [-90, 90], respectivamente. La salida de ST_GeoSquare se puede utilizar como entrada para la función ST_GeomFromGeoSquare.

Hay 360° alrededor del arco de la circunferencia ecuatorial de la Tierra que se dividen en dos hemisferios (oriental y occidental), cada uno con 180° de líneas longitudinales (meridianos) desde el meridiano 0°. Por convención, las longitudes orientales son coordenadas “+” (positivas) cuando se proyectan en un eje x sobre un plano cartesiano y las longitudes occidentales son coordenadas “-” (negativas) cuando se proyectan en un eje x sobre un plano cartesiano. Hay 90° de líneas de latitud al norte y al sur de la circunferencia ecuatorial de 0° de la Tierra, cada una de ellas paralela a dicha circunferencia. Por convención, las líneas de latitud norte intersecan el eje y "+" (positivo) cuando se proyectan en un plano cartesiano y las líneas de latitud sur intersecan el eje y "-" (negativo) cuando se proyectan en un plano cartesiano. La cuadrícula esférica formada por la intersección de las líneas longitudinales y las líneas de latitud se convierte en una cuadrícula

proyectada sobre un plano cartesiano con coordenadas “x” positivas “y” negativas estándar y coordenadas “y” positivas “y” negativas en el plano cartesiano.

El propósito de ST_GeoSquare es etiquetar o marcar puntos cercanos con valores de código iguales. Los puntos que se encuentran en el mismo geosquare reciben el mismo valor de código. Un geosquare se utiliza para codificar las coordenadas geográficas (latitud y longitud) en un número entero. Una región más amplia se divide en cuadrículas para delimitar un área en un mapa con resoluciones variables. Un geosquare se puede utilizar para la indexación espacial, la agrupación espacial, las búsquedas de proximidad, la búsqueda de ubicaciones y la creación de identificadores de lugares únicos. La función ST_GeoHash sigue un proceso similar de dividir una región en cuadrículas, pero tiene una codificación diferente.

Sintaxis

```
ST_GeoSquare(geom)
```

```
ST_GeoSquare(geom, max_depth)
```

Argumentos

geom

Un valor POINT de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo POINT. La coordenada “x” (longitud) del punto debe estar dentro del intervalo: `-180` a `180`. La coordenada “y” (latitud) del punto debe estar dentro del intervalo: `-90` a `90`.

max_depth

Un valor de tipo de datos `INTEGER`. Número máximo de veces que el dominio que contiene el punto se subdivide de forma recursiva. El valor debe ser un número entero entre 1 y 32. El valor predeterminado es 32. El número final real de las subdivisiones es inferior o igual al valor max_depth especificado.

Tipo de devolución

`BIGINT`

La función devuelve un valor único que identifica el geosquare final en el que se encuentra el punto de entrada.

Si la geometría de entrada no es un punto, la función devuelve un error.

Si el punto de entrada está vacío, el valor devuelto no es una entrada válida para la función ST_GeomFromGeoSquare. Utilice la función ST_IsEmpty para evitar las llamadas a ST_GeoSquare con un punto vacío.

Si la geometría de entrada no está dentro del intervalo, la función devuelve un error.

Si la entrada max_depth está fuera del intervalo, la función devuelve un error.

Ejemplos

El siguiente código SQL devuelve un geosquare a partir de un punto de entrada.

```
SELECT ST_GeoSquare(ST_Point(13.5, 52.5));
```

```
  st_geosquare
-----------------------
 -4410772491521635895
```

El siguiente código SQL devuelve un geosquare a partir de un punto de entrada con una profundidad máxima de 10.

```
SELECT ST_GeoSquare(ST_Point(13.5, 52.5), 10);
```

```
 st_geosquare
--------------
 797852
```

## ST_InteriorRingN

ST_InteriorRingN devuelve una cadena de línea cerrada correspondiente al anillo interior de un polígono de entrada en la posición del índice. La dimensión de la geometría devuelta es la misma que la de la geometría de entrada.

Sintaxis

```
ST_InteriorRingN(geom, index)
```

Argumentos

geom

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

índice

Un valor de tipo de datos INTEGER que representa la posición de un anillo de un índice con base uno.

Tipo de devolución

GEOMETRY del subtipo LINESTRING.

El valor del identificador de sistema de referencia espacial (SRID) de la geometría devuelta es el valor SRID de la geometría de entrada.

Si geom o index es null, entonces se devuelve el valor null.

Si índice está fuera de rango, se devuelve null.

Si geom no es un polígono, se devuelve null.

Si geom es un polígono vacío, se devuelve el valor null.

Ejemplos

El siguiente SQL devuelve el segundo anillo del polígono como una cadena de línea cerrada.

```
SELECT ST_AsEWKT(ST_InteriorRingN(ST_GeomFromText('POLYGON((7 9,8 7,11 6,15 8,16 6,17
 7,17 10,18 12,17 14,15 15,11 15,10 13,9 12,7 9),(9 9,10 10,11 11,11 10,10 8,9 9),(12
 14,15 14,13 11,12 14))'),2));
```

```
st_asewkt
-----------
 LINESTRING(12 14,15 14,13 11,12 14)
```

## ST_Intersects

ST_Intersects devuelve true si las proyecciones 2D de las dos geometrías de entrada tienen como mínimo un punto en común.

### Sintaxis

```
ST_Intersects(geom1, geom2)
```

### Argumentos

geom1

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

geom2

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

### Tipo de devolución

BOOLEAN

Si geom1 o geom2 son null, entonces se devuelve el valor null.

Si geom1 y geom2 no tienen el mismo valor en el identificador de sistema de referencia espacial (SRID), se devuelve un error.

Si geom1 o geom2 son una colección geométrica, se devuelve un error.

### Ejemplos

El siguiente SQL comprueba si el primer polígono se cruza con el segundo polígono.

```
SELECT ST_Intersects(ST_GeomFromText('POLYGON((0 0,10 0,10 10,0 10,0 0),(2 2,2 5,5 5,5
 2,2 2))'), ST_GeomFromText('MULTIPOINT((4 4),(6 6))'));
```

```
st_intersects
-------------
 true
```

## ST_Intersection

ST_Intersection devuelve una geometría que representa la intersección de conjuntos de puntos entre dos geometrías. Es decir, se devuelve la parte de las dos geometrías de entrada que se comparte entre ellas.

### Sintaxis

```
ST_Intersection(geom1, geom2)
```

### Argumentos

geom1

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

geom2

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

### Tipo de devolución

GEOMETRY

Si geom1 y geom2 no comparten ningún espacio (son discontinuos), se devuelve una geometría vacía.

Si geom1 o geom2 están vacíos, se devuelve una geometría vacía.

Si geom1 y geom2 no tienen el mismo valor en el identificador de sistema de referencia espacial (SRID), se devuelve un error.

Si geom1 o geom2 son una colección geométrica, se devuelve un error.

Si geom1 o geom2 no es una geometría bidimensional (2D), entonces se devuelve un error.

### Ejemplos

El siguiente SQL devuelve la geometría que no está vacía y representa la intersección entre dos geometrías de entrada.

```
SELECT ST_AsEWKT(ST_Intersection(ST_GeomFromText('polygon((0 0,100 100,0 200,0 0))'),
 ST_GeomFromText('polygon((0 0,10 0,0 10,0 0))')));
```

```
        st_asewkt
-------------------------
```

```
 POLYGON((0 0,0 10,5 5,0 0))
```

El siguiente SQL devuelve una geometría vacía cuando se introducen geometrías de entrada discontinuas (sin intersección).

```
SELECT ST_AsEWKT(ST_Intersection(ST_GeomFromText('linestring(0 100,0 0)'),
 ST_GeomFromText('polygon((1 0,10 0,1 10,1 0))')));
```

```
    st_asewkt
------------------
 LINESTRING EMPTY
```

## ST_IsPolygonCCW

ST_IsPolygonCCW devuelve true si la proyección 2D del multipolígono o del polígono de entrada es en sentido contrario a las agujas del reloj. Si la geometría de entrada es un punto, una cadena de línea, un multipunto o una cadena de varias líneas, se devuelve true. En el caso de las colecciones de geometría, ST_IsPolygonCCW devuelve true si todas las geometrías de la colección están en el sentido contrario a las agujas del reloj.

Sintaxis

```
ST_IsPolygonCCW(geom)
```

Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

Tipo de devolución

`BOOLEAN`

Si geom es nulo, entonces se devuelve el valor nulo.

Ejemplos

El siguiente SQL verifica si el polígono es en sentido contrario a las agujas del reloj.

```
SELECT ST_IsPolygonCCW(ST_GeomFromText('POLYGON((7 9,8 7,11 6,15 8,16 6,17 7,17 10,18
 12,17 14,15 15,11 15,10 13,9 12,7 9),(9 9,10 10,11 11,11 10,10 8,9 9),(12 14,15 14,13
 11,12 14))'));
```

```
 st_ispolygonccw
----------
 true
```

## ST_IsPolygonCW

ST_IsPolygonCW devuelve true si la proyección 2D del multipolígono o del polígono de entrada es en sentido de las agujas del reloj. Si la geometría de entrada es un punto, una cadena de línea, un multipunto o una cadena de varias líneas, se devuelve true. En el caso de las colecciones de geometría, ST_IsPolygonCW devuelve true si todas las geometrías de la colección son en el sentido de las agujas del reloj.

### Sintaxis

```
ST_IsPolygonCW(geom)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

### Tipo de devolución

`BOOLEAN`

Si geom es nulo, entonces se devuelve el valor nulo.

### Ejemplos

El siguiente SQL verifica si el polígono es en sentido de las agujas del reloj.

```
SELECT ST_IsPolygonCW(ST_GeomFromText('POLYGON((7 9,8 7,11 6,15 8,16 6,17 7,17 10,18
 12,17 14,15 15,11 15,10 13,9 12,7 9),(9 9,10 10,11 11,11 10,10 8,9 9),(12 14,15 14,13
 11,12 14))'));
```

```
 st_ispolygonccw
----------
 true
```

## ST_IsClosed

ST_IsClosed devuelve true si la proyección 2D de la geometría de entrada está cerrada. Las siguientes reglas definen una geometría cerrada:

- La geometría de entrada de datos es un punto o un multipoint.
- La geometría de entrada de datos es un linestring, y los puntos de inicio y final del linestring coinciden.
- La geometría de entrada de datos es un multilinestring que no está vacío y todos sus linestrings están cerrados.
- La geometría de entrada de datos es un polígono que no está vacío, ningún anillo del polígono está vacío, y los puntos de inicio y final de todos sus anillos coinciden.
- La geometría de entrada de datos es un multipolygon que no está vacío y todos sus polígonos están cerrados.
- La geometría de entrada de datos es una colección de geometría que no está vacía y todos sus componentes están cerrados.

### Sintaxis

```
ST_IsClosed(geom)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

### Tipo de devolución

`BOOLEAN`

Si geom es un punto vacío, se devuelve false.

Si geom es nulo, entonces se devuelve el valor nulo.

Ejemplos

El siguiente SQL comprueba si el polígono está cerrado.

```
SELECT ST_IsClosed(ST_GeomFromText('POLYGON((0 2,1 1,0 -1,0 2))'));
```

```
st_isclosed
-----------
 true
```

## ST_IsCollection

ST_IsCollection devuelve true si la geometría de entrada de datos tiene uno de los siguientes subtipos: `GEOMETRYCOLLECTION`, `MULTIPOINT`, `MULTILINESTRING` o `MULTIPOLYGON`.

Sintaxis

```
ST_IsCollection(geom)
```

Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

Tipo de devolución

`BOOLEAN`

Si geom es nulo, entonces se devuelve el valor nulo.

Ejemplos

El siguiente SQL comprueba si el polígono es una colección.

```
SELECT ST_IsCollection(ST_GeomFromText('POLYGON((0 2,1 1,0 -1,0 2))'));
```

```
st_iscollection
-----------
 false
```

## ST_IsEmpty

ST_IsEmpty devuelve true si la geometría de entrada de datos está vacía. Una geometría no está vacía si contiene al menos un punto no vacío.

ST_IsEmpty devuelve true si la geometría de entrada tiene al menos un punto no vacío.

### Sintaxis

```
ST_IsEmpty(geom)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

### Tipo de devolución

`BOOLEAN`

Si geom es nulo, entonces se devuelve el valor nulo.

### Ejemplos

El siguiente SQL comprueba si el polígono especificado está vacío.

```
SELECT ST_IsEmpty(ST_GeomFromText('POLYGON((0 2,1 1,0 -1,0 2))'));
```

```
st_isempty
-----------
 false
```

## ST_IsRing

ST_IsRing devuelve el valor verdadero si la cadena de líneas de entrada es un anillo. Una cadena de líneas es un anillo si está cerrada y es sencilla.

### Sintaxis

```
ST_IsRing(geom)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`. La geometría debe ser una `LINESTRING`.

### Tipo de devolución

`BOOLEAN`

Si geom no es `LINESTRING`, se devuelve un error.

### Ejemplos

El siguiente SQL verifica si la cadena de líneas especificada es un anillo.

```
SELECT ST_IsRing(ST_GeomFromText('linestring(0 0, 1 1, 1 2, 0 0)'));
```

```
st_isring
-----------
 true
```

## ST_IsSimple

ST_IsSimple devuelve true si la proyección 2D de la geometría de entrada es simple. Para obtener más información acerca de la definición de una geometría simple, consulte Simplicidad geométrica.

### Sintaxis

```
ST_IsSimple(geom)
```

Argumentos

geom

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

Tipo de devolución

BOOLEAN

Si geom es nulo, entonces se devuelve el valor nulo.

Ejemplos

El siguiente SQL comprueba si la cadena de línea especificada es simple. En este ejemplo, no es simple porque tiene una autointersección.

```
SELECT ST_IsSimple(ST_GeomFromText('LINESTRING(0 0,10 0,5 5,5 -5)'));
```

```
 st_issimple
-----------
 false
```

## ST_IsValid

ST_IsValid devuelve true si la proyección 2D de la geometría de entrada es válida. Para obtener más información acerca de la definición de una geometría válida, consulte Validez geométrica.

Sintaxis

```
ST_IsValid(geom)
```

Argumentos

geom

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

Tipo de devolución

`BOOLEAN`

Si geom es nulo, entonces se devuelve el valor nulo.

Ejemplos

El siguiente SQL verifica si el polígono especificado es válido. En este ejemplo, el polígono no es válido porque su interior no está conectado de forma sencilla.

```
SELECT ST_IsValid(ST_GeomFromText('POLYGON((0 0,10 0,10 10,0 10,0 0),(5 0,10 5,5 10,0
 5,5 0))'));
```

```
 st_isvalid
-----------
 false
```

## ST_Length

En el caso de una geometría lineal, ST_Length devuelve la longitud cartesiana de una proyección 2D. Las unidades de longitud son las mismas que las unidades en las que se expresan las coordenadas de la geometría de entrada. La función devuelve cero (0) para puntos, multipuntos y geometrías de área. Cuando la entrada es una colección de geometría, la función devuelve la suma de las longitudes de las geometrías de la colección.

Con respecto a una geografía, ST_Length devuelve la longitud geodésica de la proyección 2D de una geografía lineal de entrada calculada sobre el esferoide que determina el SRID. La unidad de longitud se expresa en metros. La función devuelve cero (0) para puntos, multipuntos y geografías de área. Cuando la entrada es una colección de geometría, la función devuelve la suma de las longitudes de las geografías de esa colección.

Sintaxis

```
ST_Length(geo)
```

### Argumentos

geo

Un valor de tipo de datos GEOMETRY o GEOGRAPHY, o una expresión que toma el valor de un tipo GEOMETRY o GEOGRAPHY.

### Tipo de devolución

DOUBLE PRECISION

Si geo es nulo, entonces se devuelve el valor nulo.

Si no se encuentra el valor del SRID, se devuelve un error.

### Ejemplos

El siguiente SQL devuelve la longitud cartesiana de una multilinestring.

```
SELECT ST_Length(ST_GeomFromText('MULTILINESTRING((0 0,10 0,0 10),(10 0,20 0,20
 10))'));
```

```
st_length
--------------------------------
  44.142135623731
```

El siguiente SQL devuelve la longitud de una cadena de líneas en una geografía.

```
SELECT ST_Length(ST_GeogFromText('SRID=4326;LINESTRING(5 0,6 0,4 0)'));
```

```
 st_length
------------------
 333958.472379804
```

El siguiente SQL devuelve la longitud de un punto en una geografía.

```
SELECT ST_Length(ST_GeogFromText('SRID=4326;POINT(4 5)'));
```

```
 st_length
-----------
 0
```

## ST_LengthSphere

ST_LengthSphere devuelve la longitud de una geometría lineal en metros. En el caso de las geometrías de punto, multipunto y área, ST_LengthSphere devuelve 0. En el caso de las colecciones de geometría, ST_LengthSphere devuelve la longitud total de las geometrías lineales de la colección en metros.

ST_LengthSphere interpreta las coordenadas de cada punto de la geometría de entrada como la longitud y la latitud en grados. En el caso de las geometrías 3DZ, 3DM o 4D, solo se utilizan las dos primeras coordenadas.

### Sintaxis

```
ST_LengthSphere(geom)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

### Tipo de devolución

`DOUBLE PRECISION` de longitud en metros. El cálculo de la longitud se basa en el modelo esférico de la Tierra, cuyo radio es el radio medio de la Tierra del modelo elipsoidal del Sistema Geodésico Mundial (WGS) 84 de la Tierra.

Si geom es nulo, entonces se devuelve el valor nulo.

### Ejemplos

En el siguiente ejemplo de SQL, se calcula la longitud de una cadena de línea en metros.

```
SELECT ST_LengthSphere(ST_GeomFromText('LINESTRING(10 10,45 45)'));
```

```
 st_lengthsphere
------------------
 5127736.08292556
```

## ST_Length2D

ST_Length2D es un alias para ST_Length. Para obtener más información, consulte ST_Length.

## ST_LineFromMultiPoint

ST_LineFromMultiPoint devuelve una linestring a partir de una geometría multipunto de entrada. El orden de los puntos se conserva. El identificador del sistema de referencia espacial (SRID) de la geometría devuelta es el mismo que el de la geometría de entrada. La dimensión de la geometría devuelta es la misma que la de la geometría de entrada.

Sintaxis

```
ST_LineFromMultiPoint(geom)
```

Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`. El subtipo debe ser `MULTIPOINT` .

Tipo de devolución

`GEOMETRY`

Si geom es nulo, entonces se devuelve el valor nulo.

Si geom está vacío, se devuelve una `LINESTRING` vacía.

Si geom contiene puntos vacíos, se ignoran estos puntos vacíos.

Si geom no es un `MULTIPOINT`, se devuelve error.

Ejemplos

El siguiente SQL crea una linestring a partir de un multipunto.

```
SELECT ST_AsEWKT(ST_LineFromMultiPoint(ST_GeomFromText('MULTIPOINT(0 0,10 0,10 10,5 5,0
 5)',4326)));
```

```
 st_asewkt
-------------------------------------------
 SRID=4326;LINESTRING(0 0,10 0,10 10,5 5,0 5)
```

## ST_LineInterpolatePoint

ST_LineInterpolatePoint devuelve un punto a lo largo de una línea a una distancia fraccionaria desde el inicio de la línea.

Para determinar la igualdad de los puntos, ST_LineInterpolatePoint opera en la proyección 2D de la geometría de entrada. Si la geometría de entrada está vacía, se devuelve una copia de ella en la misma dimensión que la entrada. Para las geometrías 3DZ, 3DM y 4D, la coordenada `z` o `m` es el promedio de las coordenadas `z` o `m` del segmento donde se encuentra el punto.

Sintaxis

```
ST_LineInterpolatePoint(geom, fraction)
```

Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`. El subtipo es `LINESTRING`.

fraction

Un valor de tipo de datos `DOUBLE PRECISION` que representa la posición de un punto a lo largo de la cadena de línea para la línea. El valor es una fracción en el rango de 0 a 1, inclusive.

Tipo de devolución

`GEOMETRY` del subtipo `POINT`.

Si geom o fraction son null, se devuelve el valor null.

Si geom está vacío, se devuelve el punto vacío.

El valor del identificador de sistema de referencia espacial (SRID) de la geometría devuelta es el valor SRID de la geometría de entrada.

Si fraction está fuera de rango, se devuelve un error.

Si geom no es una cadena de línea, se devuelve un error.

Ejemplos

El siguiente SQL devuelve un punto a la mitad de una cadena de línea.

```
SELECT ST_AsEWKT(ST_LineInterpolatePoint(ST_GeomFromText('LINESTRING(0 0, 5 5, 7 7, 10
 10)'), 0.50));
```

```
st_asewkt
-----------
 POINT(5 5)
```

El siguiente SQL devuelve un punto que se encuentra al 90 % del camino de una cadena de línea.

```
SELECT ST_AsEWKT(ST_LineInterpolatePoint(ST_GeomFromText('LINESTRING(0 0, 5 5, 7 7, 10
 10)'), 0.90));
```

```
st_asewkt
-----------
 POINT(9 9)
```

## ST_M

ST_M devuelve la coordenada `m` de un punto de entrada.

Sintaxis

```
ST_M(point)
```

Argumentos

point

Un valor `POINT` de tipo de datos `GEOMETRY`.

Tipo de devolución

Valor `DOUBLE PRECISION` de la coordenada `m`.

Si point es null, entonces se devuelve el valor null.

Si point es un punto 2D o 3DZ, se devuelve el valor null.

Si point es un punto vacío, se devuelve el valor null.

Si point no es `POINT`, se devuelve un error.

Ejemplos

El siguiente SQL devuelve la coordenada `m` de un punto en una geometría 3DM.

```
SELECT ST_M(ST_GeomFromEWKT('POINT M (1 2 3)'));
```

```
st_m
-----------
 3
```

El siguiente SQL devuelve la coordenada `m` de un punto en una geometría 4D.

```
SELECT ST_M(ST_GeomFromEWKT('POINT ZM (1 2 3 4)'));
```

```
st_m
-----------
 4
```

## ST_MakeEnvelope

ST_MakeEnvelope devuelve una geometría de la siguiente manera:

- Si las coordenadas de entrada especifican un punto, la geometría devuelta es un punto.
- Si las coordenadas de entrada especifican una línea, la geometría devuelta es una cadena de línea.

- De lo contrario, la geometría devuelta es un polígono, donde las coordenadas de entrada especifican las esquinas inferior izquierda y superior derecha de un cuadro.

Si se proporciona, el valor del identificador de sistema de referencia espacial (SRID) de la geometría devuelta se establece en el valor SRID de entrada.

Sintaxis

```
ST_MakeEnvelope(xmin, ymin, xmax, ymax)
```

```
ST_MakeEnvelope(xmin, ymin, xmax, ymax, srid)
```

Argumentos

xmin

Un valor de tipo de datos `DOUBLE PRECISION`. Este valor es la primera coordenada de la esquina inferior izquierda de un cuadro.

ymin

Un valor de tipo de datos `DOUBLE PRECISION`. Este valor es la segunda coordenada de la esquina inferior izquierda de un cuadro.

xmax

Un valor de tipo de datos `DOUBLE PRECISION`. Este valor es la primera coordenada de la esquina superior derecha de un cuadro.

ymax

Un valor de tipo de datos `DOUBLE PRECISION`. Este valor es la segunda coordenada de la esquina superior derecha de un cuadro.

srid

Un valor de tipo de datos `INTEGER` que representa un identificador de referencia espacial (SRID). Si no se proporciona el valor de SRID, se establece en cero.

Tipo de devolución

`GEOMETRY` del subtipo `POINT`, `LINESTRING` o `POLYGON`.

El SRID de la geometría devuelta se establece en `srid` o cero si `srid` no está configurado.

Si xmin, ymin, xmax, ymax o srid es null, se devolverá un valor nulo.

Si srid es negativo, se devuelve un error.

Ejemplos

El siguiente SQL devuelve un polígono que representa un sobre definido por los cuatro valores de las coordenadas de entrada.

```
SELECT ST_AsEWKT(ST_MakeEnvelope(2,4,5,7));
```

```
 st_astext
---------------
 POLYGON((2 4,2 7,5 7,5 4,2 4))
```

El siguiente SQL devuelve un polígono que representa un sobre definido por los cuatro valores de las coordenadas de entrada y un valor SRID.

```
SELECT ST_AsEWKT(ST_MakeEnvelope(2,4,5,7,4326));
```

```
 st_astext
---------------------------------
 SRID=4326;POLYGON((2 4,2 7,5 7,5 4,2 4))
```

## ST_MakeLine

ST_MakeLine crea un linestring a partir de las geometrías de entrada de datos.

La dimensión de la geometría devuelta es la misma que la de las geometrías de entrada. Ambas geometrías de entrada deben ser de la misma dimensión.

Sintaxis

```
ST_MakeLine(geom1, geom2)
```

Argumentos

geom1

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY. El subtipo debe ser POINT, LINESTRING o MULTIPOINT.

geom2

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY. El subtipo debe ser POINT, LINESTRING o MULTIPOINT.

Tipo de devolución

GEOMETRY del subtipo LINESTRING.

Si geom1 o geom2 son null, entonces se devuelve el valor null.

Si geom1 y geom2 son el punto vacío o contiene puntos vacíos, se ignoran estos puntos vacíos.

Si geom1 y geom2 están vacíos, se devuelve la LINESTRING vacía.

El valor del identificador de sistema de referencia espacial (SRID) de la geometría devuelta es el valor SRID de las geometrías de entrada de datos.

Si geom1 y geom2 tienen diferentes valores SRID, se devuelve un error.

Si geom1 o geom2 no es un elemento POINT, LINESTRING o MULTIPOINT, se devuelve un error.

Si geom1 y geom2 tienen diferentes dimensiones, se devuelve un error.

Ejemplos

El siguiente SQL construye un linestring a partir de dos linestring de entrada de datos.

```
SELECT ST_MakeLine(ST_GeomFromText('LINESTRING(77.29 29.07,77.42 29.26,77.27
 29.31,77.29 29.07)'), ST_GeomFromText('LINESTRING(88.29 39.07,88.42 39.26,88.27
 39.31,88.29 39.07)'));
```

```
st_makeline
-----------
```

```
0102000000080000000C3F5285C8F52534052B81E85EB113D407B14AE47E15A5340C3F5285C8F423D40E17A14AE4751
```

## ST_MakePoint

ST_MakePoint devuelve una geometría de punto los valores de cuyas coordenadas son los valores de entrada.

### Sintaxis

```
ST_MakePoint(x, y)
```

```
ST_MakePoint(x, y, z)
```

```
ST_MakePoint(x, y, z, m)
```

### Argumentos

x

Un valor de tipo de datos `DOUBLE PRECISION` representa la primera coordenada.

y

Un valor de tipo de datos `DOUBLE PRECISION` representa la segunda coordenada.

z

Un valor de tipo de datos `DOUBLE PRECISION` que representa la tercera coordenada.

m

Un valor de tipo de datos `DOUBLE PRECISION` que representa la cuarta coordenada.

### Tipo de devolución

`GEOMETRY` del subtipo `POINT`.

El valor del identificador de sistema de referencia espacial (SRID) de la geometría devuelta se establece en 0.

Si x, y, z o m son nulos, se devuelve un valor nulo.

### Ejemplos

El siguiente SQL devuelve un tipo GEOMETRY de subtipo POINT con las coordinadas proporcionadas.

```
SELECT ST_AsText(ST_MakePoint(1,3));
```

```
st_astext
-----------
 POINT(1 3)
```

El siguiente SQL devuelve un tipo GEOMETRY de subtipo POINT con las coordinadas proporcionadas.

```
SELECT ST_AsEWKT(ST_MakePoint(1, 2, 3));
```

```
st_asewkt
----------------
 POINT Z (1 2 3)
```

El siguiente SQL devuelve un tipo GEOMETRY de subtipo POINT con las coordinadas proporcionadas.

```
SELECT ST_AsEWKT(ST_MakePoint(1, 2, 3, 4));
```

```
st_asewkt
-------------------
 POINT ZM (1 2 3 4)
```

## ST_MakePolygon

ST_MakePolygon tiene dos variantes que devuelven un polígono. Una toma una sola geometría y la otra toma dos geometrías.

- La entrada de la primera variante es una cadena de línea que define el anillo exterior del polígono de salida.
- La entrada de la segunda variante es una cadena de línea y una cadena de varias líneas. Ambas están vacías o cerradas.

El límite del anillo exterior del polígono de salida es la cadena de línea de entrada y los límites de los anillos interiores del polígono son las cadenas de líneas en la cadena de varias líneas de entrada. Si la cadena de línea de entrada está vacía, se devuelve un polígono vacío. Las cadenas de líneas vacías en la cadena de varias líneas no se tienen en cuenta. El identificador de sistema de referencia espacial (SRID) de la geometría resultante es el SRID en común de las geometrías de entrada.

La dimensión de la geometría devuelta es la misma que la de las geometrías de entrada. El anillo exterior y los anillos interiores deben tener la misma dimensión.

Sintaxis

```
ST_MakePolygon(geom1)
```

```
ST_MakePolygon(geom1, geom2)
```

Argumentos

geom1

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY. El subtipo debe ser LINESTRING . El valor linestring debe ser cerrado o estar vacío.

geom2

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY. El subtipo debe ser MULTILINESTRING .

Tipo de devolución

GEOMETRY del subtipo POLYGON.

El identificador de sistema de referencia espacial (SRID) de la geometría devuelta es igual al SRID de las entradas.

Si geom1 o geom2 son null, se devuelve el valor null.

Si geom1 no es una cadena de línea, se devuelve un error.

Si geom2 no es una cadena de varias líneas, se devuelve un error.

Si geom1 no está cerrado, se devuelve un error.

Si geom1 es un punto único o no está cerrado, se devuelve un error.

Si geom2 contiene al menos una cadena de línea que tiene un punto único o no está cerrada, se devuelve un error.

Si geom1 y geom2 tienen diferentes valores SRID, se devuelve un error.

Si geom1 y geom2 tienen diferentes dimensiones, se devuelve un error.

Ejemplos

El siguiente SQL devuelve un polígono a partir de un linestring de entrada de datos.

```
SELECT ST_AsText(ST_MakePolygon(ST_GeomFromText('LINESTRING(77.29 29.07,77.42
 29.26,77.27 29.31,77.29 29.07)')));
```

```
 st_astext
---------------
POLYGON((77.29 29.07,77.42 29.26,77.27 29.31,77.29 29.07))
```

El siguiente SQL crea un polígono a partir de una cadena de línea cerrada y una cadena de varias líneas cerrada. La cadena de línea se utiliza para el anillo exterior del polígono. Las cadenas de líneas en las cadenas de varias líneas se utilizan para los anillos interiores del polígono.

```
SELECT ST_AsEWKT(ST_MakePolygon(ST_GeomFromText('LINESTRING(0 0,10 0,10 10,0 10,0 0)'),
 ST_GeomFromText('MULTILINESTRING((1 1,1 2,2 1,1 1),(3 3,3 4,4 3,3 3))')));
```

```
 st_astext
---------------------------------
POLYGON((0 0,10 0,10 10,0 10,0 0),(1 1,1 2,2 1,1 1),(3 3,3 4,4 3,3 3))
```

## ST_MemSize

ST_MemSize devuelve la cantidad de espacio de memoria (en bytes) utilizado por la geometría de entrada. Este tamaño depende de la representación interna de Amazon Redshift de la geometría y,

por lo tanto, puede cambiar si la representación interna cambia. Puede utilizar este tamaño como una indicación del tamaño relativo de los objetos de geometría en Amazon Redshift.

Sintaxis

```
ST_MemSize(geom)
```

Argumentos

geom

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

Tipo de devolución

INTEGER representa la dimensión inherente de geom.

Si geom es nulo, entonces se devuelve el valor nulo.

Ejemplos

El siguiente SQL devuelve el tamaño de memoria de una colección de geometría.

```
SELECT ST_MemSize(ST_GeomFromText('GEOMETRYCOLLECTION(POLYGON((0 0,10 0,0 10,0
 0)),LINESTRING(20 10,20 0,10 0))')::varchar + ' bytes';
```

```
 ?column?
-----------
 172 bytes
```

## ST_MMax

ST_MMax devuelve la coordenada m máxima de una geometría de entrada.

Sintaxis

```
ST_MMax(geom)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

### Tipo de devolución

Valor `DOUBLE PRECISION` de la primera coordenada `m` máxima.

Si geom está vacío, entonces se devuelve el valor null.

Si geom es nulo, entonces se devuelve el valor nulo.

Si geom es una geometría 2D o 3DZ, se devuelve el valor null.

### Ejemplos

El siguiente SQL devuelve la coordenada `m` más grande de una cadena de línea en una geometría 3DM.

```
SELECT ST_MMax(ST_GeomFromEWKT('LINESTRING M (0 1 2, 3 4 5, 6 7 8)'));
```

```
st_mmax
-----------
  8
```

El siguiente SQL devuelve la coordenada `m` más grande de una cadena de línea en una geometría 4D.

```
SELECT ST_MMax(ST_GeomFromEWKT('LINESTRING ZM (0 1 2 3, 4 5 6 7, 8 9 10 11)'));
```

```
st_mmax
-----------
  11
```

## ST_MMin

ST_MMin devuelve la coordenada `m` mínima de una geometría de entrada.

### Sintaxis

```
ST_MMin(geom)
```

### Argumentos

geom

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

### Tipo de devolución

Valor DOUBLE PRECISION de la coordenada m mínima.

Si geom está vacío, entonces se devuelve el valor null.

Si geom es nulo, entonces se devuelve el valor nulo.

Si geom es una geometría 2D o 3DZ, se devuelve el valor null.

### Ejemplos

El siguiente SQL devuelve la coordenada m más pequeña de una cadena de línea en una geometría 3DM.

```
SELECT ST_MMin(ST_GeomFromEWKT('LINESTRING M (0 1 2, 3 4 5, 6 7 8)'));
```

```
st_mmin
-----------
  2
```

El siguiente SQL devuelve la coordenada m más pequeña de una cadena de línea en una geometría 4D.

```
SELECT ST_MMin(ST_GeomFromEWKT('LINESTRING ZM (0 1 2 3, 4 5 6 7, 8 9 10 11)'));
```

```
st_mmin
-----------
```

```
   3
```

## ST_Multi

ST_Multi convierte la geometría en el multitipo correspondiente. Si la geometría de entrada ya es un multitipo o una colección de geometría, se devuelve una copia de ella. Si la geometría de entrada es un punto, una cadena de línea o un polígono, se devuelve un multipunto, una cadena de varias líneas o un multipolígono, respectivamente, que contiene la geometría de entrada.

### Sintaxis

```
ST_Multi(geom)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

### Tipo de devolución

`GEOMETRY` con subtipo `MULTIPOINT`, `MULTILINESTRING`, `MULTIPOLYGON` o `GEOMETRYCOLLECTION`.

El identificador del sistema de referencia espacial (SRID) de la geometría devuelta es el mismo que el de la geometría de entrada.

Si geom es nulo, entonces se devuelve el valor nulo.

### Ejemplos

El siguiente SQL devuelve un multipunto a partir de un multipunto de entrada.

```
SELECT ST_AsEWKT(ST_Multi(ST_GeomFromText('MULTIPOINT((1 2),(3 4))', 4326)));
```

```
    st_asewkt
-----------------------------------
  SRID=4326;MULTIPOINT((1 2),(3 4))
```

El siguiente SQL devuelve un multipunto a partir de un punto de entrada.

```
SELECT ST_AsEWKT(ST_Multi(ST_GeomFromText('POINT(1 2)', 4326)));
```

```
    st_asewkt
-----------------------------------
  SRID=4326;MULTIPOINT((1 2))
```

El siguiente SQL devuelve una colección de geometría a partir de una colección de geometría de entrada.

```
SELECT ST_AsEWKT(ST_Multi(ST_GeomFromText('GEOMETRYCOLLECTION(POINT(1 2),MULTIPOINT((1
 2),(3 4)))', 4326)));
```

```
    st_asewkt
-----------------------------------
  SRID=4326;GEOMETRYCOLLECTION(POINT(1 2),MULTIPOINT((1 2),(3 4)))
```

## ST_NDims

ST_NDims devuelve la dimensión de las coordenadas de una geometría. ST_NDims no tiene en cuenta la dimensión topológica de una geometría. En su lugar, devuelve un valor constante en función de la dimensión de la geometría.

### Sintaxis

```
ST_NDims(geom)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

### Tipo de devolución

`INTEGER` representa la dimensión inherente de geom.

Si geom es nulo, entonces se devuelve el valor nulo.

Los valores retornados son los siguientes.

| Valor devuelto | Dimensión de la geometría de entrada |
|---|---|
| 2 | 2D |
| 3 | 3DZ o 3DM |
| 4 | 4D |

### Ejemplos

El siguiente SQL devuelve la cantidad de dimensiones de una cadena de línea 2D.

```
SELECT ST_NDims(ST_GeomFromText('LINESTRING(0 0,1 1,2 2,0 0)'));
```

```
st_ndims
-------------
 2
```

El siguiente SQL devuelve la cantidad de dimensiones de una cadena de línea 3DZ.

```
SELECT ST_NDims(ST_GeomFromText('LINESTRING Z(0 0 3,1 1 3,2 2 3,0 0 3)'));
```

```
st_ndims
-------------
 3
```

El siguiente SQL devuelve la cantidad de dimensiones de una cadena de línea 3DM.

```
SELECT ST_NDims(ST_GeomFromText('LINESTRING M(0 0 4,1 1 4,2 2 4,0 0 4)'));
```

```
st_ndims
-------------
 3
```

El siguiente SQL devuelve la cantidad de dimensiones de una cadena de línea 4D.

```
SELECT ST_NDims(ST_GeomFromText('LINESTRING ZM(0 0 3 4,1 1 3 4,2 2 3 4,0 0 3 4)'));
```

```
st_ndims
-------------
 4
```

## ST_NPoints

ST_NPoints devuelve el número de puntos que no están vacíos en una geometría o geografía de entrada.

### Sintaxis

```
ST_NPoints(geo)
```

### Argumentos

geo

Un valor de tipo de datos GEOMETRY o GEOGRAPHY, o una expresión que toma el valor de un tipo GEOMETRY o GEOGRAPHY.

### Tipo de devolución

INTEGER

Si geo es un punto vacío, se devuelve 0.

Si geo es nulo, entonces se devuelve el valor nulo.

### Ejemplos

El siguiente SQL devuelve el número de puntos en un linestring.

```
SELECT ST_NPoints(ST_GeomFromText('LINESTRING(77.29 29.07,77.42 29.26,77.27 29.31,77.29
 29.07)'));
```

```
st_npoints
-------------
 4
```

El siguiente SQL devuelve el número de puntos de una cadena de líneas en una geografía.

```
SELECT ST_NPoints(ST_GeogFromText('LINESTRING(110 40, 2 3, -10 80, -7 9)'));
```

```
st_npoints
-------------
 4
```

## ST_NRings

ST_NRings devuelve el número de anillos en una geometría de entrada de datos.

Sintaxis

```
ST_NRings(geom)
```

Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

Tipo de devolución

`INTEGER`

Si geom es nulo, entonces se devuelve el valor nulo.

Los valores retornados son los siguientes.

| Valor devuelto | Subtipo de geometría. |
| --- | --- |
| 0 | Se devuelve si geom es un subtipo de `POINT`, `LINESTRING` , `MULTIPOINT` o `MULTILINESTRING` |

| Valor devuelto | Subtipo de geometría. |
| --- | --- |
| El número de anillos. | Se devuelve si geom es un subtipo de POLYGON o MULTIPOLYGON |
| El número de anillos en todos sus component es | Se devuelve si geom es un subtipo de GEOMETRYCOLLECTION |

Ejemplos

El siguiente SQL devuelve el número de anillos en un multipolygon.

```
SELECT ST_NRings(ST_GeomFromText('MULTIPOLYGON(((0 0,10 0,0 10,0 0)),((0 0,-10 0,0
 -10,0 0)))'));
```

```
 st_nrings
-------------
 2
```

## ST_NumGeometries

ST_NumGeometries devuelve el número de geometrías que hay en una geometría de entrada de datos.

Sintaxis

```
ST_NumGeometries(geom)
```

Argumentos

geom

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

Tipo de devolución

INTEGER representa el número de geometrías en geom.

Si geom es nulo, entonces se devuelve el valor nulo.

Si geom es una geometría única y vacía, se devuelve `0`.

Si geom es una geometría única y no vacía, se devuelve 1.

Si geom es un subtipo de `GEOMETRYCOLLECTION` o `MULTI`, se devuelve la cantidad de geometrías.

Ejemplos

El siguiente SQL devuelve el número de geometrías en el multilinestring de entrada de datos.

```
SELECT ST_NumGeometries(ST_GeomFromText('MULTILINESTRING((0 0,1 0,0 5),(3 4,13 26))'));
```

```
st_numgeometries
-------------
 2
```

## ST_NumInteriorRings

ST_NumInteriorRings devuelve el número de anillos en una geometría de polígono de entrada de datos.

Sintaxis

```
ST_NumInteriorRings(geom)
```

Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

Tipo de devolución

`INTEGER`

Si geom es nulo, entonces se devuelve el valor nulo.

Si geom no es un polígono, se devuelve null.

### Ejemplos

El siguiente SQL devuelve el número de anillos internos en el polígono de entrada de datos.

```
SELECT ST_NumInteriorRings(ST_GeomFromText('POLYGON((0 0,100 0,100 100,0 100,0 0),(1
 1,1 5,5 1,1 1),(7 7,7 8,8 7,7 7))'));
```

```
 st_numinteriorrings
-------------
 2
```

## ST_NumPoints

ST_NumPoints devuelve el número de puntos en una geometría de entrada de datos.

### Sintaxis

```
ST_NumPoints(geom)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

### Tipo de devolución

`INTEGER`

Si geom es nulo, entonces se devuelve el valor nulo.

Si geom no es de subtipo `LINESTRING`, entonces se devuelve null.

### Ejemplos

El siguiente SQL devuelve el número de puntos en un linestring de entrada de datos.

```
SELECT ST_NumPoints(ST_GeomFromText('LINESTRING(77.29 29.07,77.42 29.26,77.27
 29.31,77.29 29.07)'));
```

```
st_numpoints
-------------
4
```

El siguiente SQL devuelve null porque el geom de entrada no es de subtipo `LINESTRING`.

```
SELECT ST_NumPoints(ST_GeomFromText('MULTIPOINT(1 2,3 4)'));
```

```
st_numpoints
-------------

```

## ST_Perimeter

Con respecto a una geometría de área de entrada, ST_Perimeter devuelve el perímetro cartesiano (longitud del límite) de la proyección 2D. Las unidades perimetrales son las mismas que las unidades en las que se expresan las coordenadas de la geometría de entrada. La función devuelve cero (0) para geometrías de puntos, multipuntos y lineales. Cuando la entrada es una colección de geometría, la función devuelve la suma de los perímetros de las geometrías de la colección.

Con respecto a una geografía de entrada, ST_Perimeter devuelve el perímetro geodésico (longitud del límite) de la proyección 2D de una geografía de área de entrada calculada sobre el esferoide que determina el SRID. La unidad de perímetro se expresa en metros. La función devuelve un valor cero (0) para las geografías de puntos, multipuntos y lineales. Cuando la entrada es una colección de geometría, la función devuelve la suma de los perímetros de las geografías de la colección.

Sintaxis

```
ST_Perimeter(geo)
```

Argumentos

geo

Un valor de tipo de datos `GEOMETRY` o `GEOGRAPHY`, o una expresión que toma el valor de un tipo `GEOMETRY` o `GEOGRAPHY`.

Tipo de devolución

`DOUBLE PRECISION`

Si geo es nulo, entonces se devuelve el valor nulo.

Si no se encuentra el valor del SRID, se devuelve un error.

Ejemplos

El siguiente SQL devuelve el perímetro cartesiano de un multipolígono.

```
SELECT ST_Perimeter(ST_GeomFromText('MULTIPOLYGON(((0 0,10 0,0 10,0 0)),((10 0,20 0,20
 10,10 0)))'));
```

```
 st_perimeter
--------------------------------
    68.2842712474619
```

El siguiente SQL devuelve el perímetro cartesiano de un multipolígono.

```
SELECT ST_Perimeter(ST_GeomFromText('MULTIPOLYGON(((0 0,10 0,0 10,0 0)),((10 0,20 0,20
 10,10 0)))'));
```

```
 st_perimeter
--------------------------------
    68.2842712474619
```

El siguiente SQL devuelve el perímetro de un polígono en una geografía.

```
SELECT ST_Perimeter(ST_GeogFromText('SRID=4326;POLYGON((0 0,1 0,0 1,0 0))'));
```

```
 st_perimeter
------------------
 378790.428393693
```

El siguiente SQL devuelve el perímetro de una cadena de líneas en una geografía.

```
SELECT ST_Perimeter(ST_GeogFromText('SRID=4326;LINESTRING(5 0,10 0)'));
```

```
 st_perimeter
--------------
 0
```

## ST_Perimeter2D

ST_Perimeter2D es un alias para ST_Perimeter. Para obtener más información, consulte ST_Perimeter.

## ST_Point

ST_Point devuelve una geometría de punto a partir de los valores de las coordenadas de entrada de datos.

Sintaxis

```
ST_Point(x, y)
```

Argumentos

x

Un valor de tipo de datos `DOUBLE PRECISION` que representa una primera coordenada.

y

Un valor de tipo de datos `DOUBLE PRECISION` que representa una segunda coordenada.

Tipo de devolución

`GEOMETRY` del subtipo `POINT`.

El valor del identificador de sistema de referencia espacial (SRID) de la geometría devuelta se establece en 0.

Si x o y son null, entonces se devuelve el valor null.

Ejemplos

El siguiente SQL construye una geometría de punto a partir de las coordenadas de entrada de datos.

```
SELECT ST_AsText(ST_Point(5.0, 7.0));
```

```
st_astext
-------------
POINT(5 7)
```

## ST_PointN

ST_PointN devuelve un punto en una linestring según lo especificado por un valor de índice. Los valores de índice negativos se cuentan hacia atrás desde el final de la linestring, de modo que -1 es el último punto.

La dimensión de la geometría devuelta es la misma que la de la geometría de entrada.

Sintaxis

```
ST_PointN(geom, index)
```

Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`. El subtipo debe ser `LINESTRING` .

índice

Un valor de tipo de datos `INTEGER` que representa el índice de un punto en una linestring.

Tipo de devolución

`GEOMETRY` del subtipo `POINT`.

El valor del identificador de sistema de referencia espacial (SRID) de la geometría devuelta se establece en 0.

Si geom o index es null, entonces se devuelve el valor null.

Si índice está fuera de rango, se devuelve null.

Si geom está vacío, entonces se devuelve el valor null.

Si geom no es una `LINESTRING`, entonces se devuelve el valor null.

Ejemplos

El siguiente SQL devuelve una representación extended well-known text (EWKT) de una `LINESTRING` de seis puntos a un objeto `GEOMETRY` y devuelve el punto en el índice 5 de la linestring.

```
SELECT ST_AsEWKT(ST_PointN(ST_GeomFromText('LINESTRING(0 0,10 0,10 10,5 5,0 5,0
 0)',4326), 5));
```

```
st_asewkt
-------------
 SRID=4326;POINT(0 5)
```

## ST_Points

ST_Points devuelve una geometría multipunto que contiene todos los puntos no vacíos de la geometría de entrada. ST_Points no elimina los puntos duplicados en la entrada, incluidos los puntos de inicio y de finalización de las geometrías de anillo.

Sintaxis

```
ST_Points(geom)
```

Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

Tipo de devolución

`GEOMETRY` del subtipo `MULTIPOINT`.

El valor del identificador de sistema de referencia espacial (SRID) de la geometría devuelta es el mismo que el de geom.

Si geom es nulo, entonces se devuelve el valor nulo.

Si geom está vacío, se devuelve el multipunto vacío.

Ejemplos

En los siguientes ejemplos SQL, se construye una geometría multipunto a partir de la geometría de entrada. Como resultado, se obtiene una geometría multipunto que contiene todos los puntos no vacíos de la geometría de entrada.

```
SELECT ST_AsEWKT(ST_Points(ST_SetSRID(ST_GeomFromText('LINESTRING(1 0,2 0,3 0)'),
 4326)));
```

```
st_asewkt
-------------
SRID=4326;MULTIPOINT((1 0),(2 0),(3 0))
```

```
SELECT ST_AsEWKT(ST_Points(ST_SetSRID(ST_GeomFromText('MULTIPOLYGON(((0 0,1 0,0 1,0
 0)))'), 4326)));
```

```
st_asewkt
-------------
SRID=4326;MULTIPOINT((0 0),(1 0),(0 1),(0 0))
```

## ST_Polygon

ST_Polygon devuelve una geometría de polígonos cuyo anillo exterior es el linestring de entrada de datos con el valor que era el valor de entrada del identificador del sistema de referencia espacial (SRID).

La dimensión de la geometría devuelta es la misma que la de la geometría de entrada.

### Sintaxis

```
ST_Polygon(linestring, srid)
```

### Argumentos

linestring

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY. El subtipo debe ser LINESTRING que representa un linestring. El valor linestring debe ser cerrado.

srid

Un valor de tipo de datos INTEGER que representa un SRID.

### Tipo de devolución

GEOMETRY del subtipo POLYGON.

El valor SRID de la geometría devuelta se establece en srid.

Si linestring o srid son null, entonces se devuelve el valor null.

Si linestring no es un linestring, se devuelve un error.

Si linestring no está cerrado, se devuelve un error.

Si srid es negativo, se devuelve un error.

### Ejemplos

El siguiente SQL construye un polígono con un valor SRID.

```
SELECT ST_AsEWKT(ST_Polygon(ST_GeomFromText('LINESTRING(77.29 29.07,77.42 29.26,77.27
 29.31,77.29 29.07)'),4356));
```

```
st_asewkt
-------------
 SRID=4356;POLYGON((77.29 29.07,77.42 29.26,77.27 29.31,77.29 29.07))
```

## ST_RemovePoint

ST_RemovePoint devuelve una geometría de linestring que tiene eliminado el punto de la geometría de entrada en una posición de índice.

El índice tiene base cero. El identificador del sistema de referencia espacial (SRID) del resultado es el mismo que la geometría de entrada. La dimensión de la geometría devuelta es la misma que la de la geometría de entrada.

### Sintaxis

```
ST_RemovePoint(geom, index)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`. El subtipo debe ser `LINESTRING` .

índice

Un valor del tipo de datos `INTEGER` que representa la posición de un índice con base cero.

### Tipo de devolución

`GEOMETRY`

Si geom o index es null, entonces se devuelve el valor null.

Si geom no es subtipo `LINESTRING`, entonces se devuelve un error.

Si el índice está fuera de rango, se devuelve un error. Los valores válidos para la posición de índice están entre 0 y `ST_NumPoints(geom)` menos 1.

### Ejemplos

El siguiente SQL elimina el último punto de una linestring.

```
WITH tmp(g) AS (SELECT ST_GeomFromText('LINESTRING(0 0,10 0,10 10,5 5,0 5)',4326))
SELECT ST_AsEWKT(ST_RemovePoint(g, ST_NumPoints(g) - 1)) FROM tmp;
```

```
   st_asewkt
----------------------------------------
 SRID=4326;LINESTRING(0 0,10 0,10 10,5 5)
```

## ST_Reverse

ST_Reverse invierte el orden de los vértices para las geometrías lineales y de área. Para las geometrías de puntos o multipunto, se devuelve una copia de la geometría original. En las colecciones de geometría, ST_Reverse invierte el orden de los vértices de cada una de las geometrías de la colección.

La dimensión de la geometría devuelta es la misma que la de la geometría de entrada.

### Sintaxis

```
ST_Reverse(geom)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

### Tipo de devolución

`GEOMETRY`

El identificador del sistema de referencia espacial (SRID) de la geometría devuelta es el mismo que el de la geometría de entrada.

Si geom es nulo, entonces se devuelve el valor nulo.

### Ejemplos

El siguiente SQL invierte el orden de los puntos de una cadena de línea.

```
SELECT ST_AsEWKT(ST_Reverse(ST_GeomFromText('LINESTRING(1 0,2 0,3 0,4 0)', 4326)));
```

```
    st_asewkt
-----------------------------------
  SRID=4326;LINESTRING(4 0,3 0,2 0,1 0)
```

## ST_SetPoint

ST_SetPoint devuelve una cadena de línea con coordenadas actualizadas con respecto a la posición de la cadena de línea de entrada especificada por el índice. Las coordenadas nuevas son las coordenadas del punto de entrada.

La dimensión de la geometría devuelta es la misma que la del valor geom1. Si geom1 y geom2 tienen diferentes dimensiones, geom2 se proyecta a la dimensión de geom1.

### Sintaxis

```
ST_SetPoint(geom1, index, geom2)
```

### Argumentos

geom1

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`. El subtipo debe ser `LINESTRING` .

índice

Un valor del tipo de datos `INTEGER` que representa la posición de un índice. `0` hace referencia al primer punto de la cadena de línea desde la izquierda, `1` hace referencia al segundo punto, y así sucesivamente. El índice puede ser un valor negativo. `-1` hace referencia al primer punto de la cadena de línea desde la derecha, `-2` hace referencia al segundo punto, y así sucesivamente.

geom2

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`. El subtipo debe ser `POINT` .

### Tipo de devolución

`GEOMETRY`

Si geom2 es el punto vacío, se devuelve geom1.

Si geom1, geom2, o índice es null, se devuelve el valor null.

Si geom1 no es una cadena de línea, se devuelve un error.

Si index no está dentro de un rango de índice válido, se devuelve un error.

Si geom2 no es un punto, se devuelve un error.

Si geom1 y geom2 no tienen el mismo valor en el identificador de sistema de referencia espacial (SRID), se devuelve un error.

Ejemplos

El siguiente SQL devuelve una nueva cadena de línea en la que se establece el segundo punto de la cadena de línea de entrada con el punto especificado.

```
SELECT ST_AsText(ST_SetPoint(ST_GeomFromText('LINESTRING(1 2, 3 2, 5 2, 1 2)'), 2,
 ST_GeomFromText('POINT(7 9)')));
```

```
st_astext
-------------
 LINESTRING(1 2,3 2,7 9,1 2)
```

El siguiente ejemplo de SQL devuelve una nueva cadena de línea en la que se establece el tercer punto desde la derecha (el índice es negativo) de la cadena de línea con el punto especificado.

```
SELECT ST_AsText(ST_SetPoint(ST_GeomFromText('LINESTRING(1 2, 3 2, 5 2, 1 2)'), -3,
 ST_GeomFromText('POINT(7 9)')));
```

```
st_astext
-------------
 LINESTRING(1 2,7 9,5 2,1 2)
```

## ST_SetSRID

ST_SetSRID devuelve una geometría que es la misma que la geometría de entrada de datos, pero está actualizada con el valor de entrada de datos del identificador del sistema de referencia espacial (SRID).

### Sintaxis

```
ST_SetSRID(geom, srid)
```

### Argumentos

geom

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

srid

Un valor de tipo de datos INTEGER que representa un SRID.

### Tipo de devolución

GEOMETRY

El valor SRID de la geometría devuelta se establece en srid.

Si geom o srid son null, entonces se devuelve el valor null.

Si srid es negativo, se devuelve un error.

### Ejemplos

El siguiente SQL establece el valor SRID de un linestring.

```
SELECT ST_AsEWKT(ST_SetSRID(ST_GeomFromText('LINESTRING(77.29 29.07,77.42 29.26,77.27
 29.31,77.29 29.07)'),50));
```

```
st_asewkt
-------------
 SRID=50;LINESTRING(77.29 29.07,77.42 29.26,77.27 29.31,77.29 29.07)
```

## ST_Simplify

ST_Simplify devuelve una copia simplificada de la geometría de entrada mediante el algoritmo de Ramer-Douglas-Peucker con la tolerancia dada. Es posible que no se conserve la topología de la geometría de entrada. Para obtener más información acerca del algoritmo, consulte Algoritmo de Ramer–Douglas–Peucker en Wikipedia.

Cuando ST_Simplify calcula distancias para simplificar una geometría, ST_Simplify opera en la proyección 2D de la geometría de entrada.

Sintaxis

```
ST_Simplify(geom, tolerance)
```

Argumentos

geom

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

tolerance

Un valor de tipo de datos DOUBLE PRECISION que representa el nivel de tolerancia del algoritmo de Ramer-Douglas-Peucker. Si tolerance es un número negativo, entonces se utiliza el valor cero.

Tipo de devolución

GEOMETRY.

El valor del identificador de sistema de referencia espacial (SRID) de la geometría devuelta es el valor SRID de la geometría de entrada.

La dimensión de la geometría devuelta es la misma que la de la geometría de entrada.

Si geom es nulo, entonces se devuelve el valor nulo.

Ejemplos

El siguiente SQL simplifica la cadena de línea de entrada mediante una tolerancia de distancia euclidiana de 1 con el algoritmo de Ramer-Douglas-Peucker. Las unidades de la distancia son las mismas que en las coordenadas de la geometría.

```
SELECT ST_AsEWKT(ST_Simplify(ST_GeomFromText('LINESTRING(0 0,1 2,1 1,2 2,2 1)'), 1));
```

```
 st_asewkt
-----------
LINESTRING(0 0,1 2,2 1)
```

## ST_SRID

ST_SRID devuelve el identificador del sistema de referencia espacial (SRID) de una geometría de entrada. Para obtener más información sobre SRID, consulte Consulta de datos espaciales en Amazon Redshift.

### Sintaxis

```
ST_SRID(geom)
```

### Argumentos

geom

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

### Tipo de devolución

INTEGER representa el valor SRID de geom.

Si geom es nulo, entonces se devuelve el valor nulo.

### Ejemplos

El siguiente SQL devuelve un valor SRID de una cadena de línea que se establece en SRID 4326.

```
SELECT ST_SRID(ST_GeomFromText('LINESTRING(77.29 29.07,77.42 29.26,77.27 29.31,77.29
 29.07)',4326));
```

```
st_srid
-------------
 4326
```

El siguiente SQL devuelve un valor SRID de una cadena de línea que no está establecida cuando se construye. Esto resulta en 0 para el valor SRID.

```
SELECT ST_SRID(ST_GeomFromText('LINESTRING(77.29 29.07,77.42 29.26,77.27 29.31,77.29
 29.07)'));
```

```
st_srid
-------------
 0
```

## ST_StartPoint

ST_StartPoint devuelve el primer punto de una linestring de entrada. El valor del identificador del sistema de referencia espacial (SRID) del resultado es el mismo que el de la geometría de entrada. La dimensión de la geometría devuelta es la misma que la de la geometría de entrada.

### Sintaxis

```
ST_StartPoint(geom)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`. El subtipo debe ser `LINESTRING` .

### Tipo de devolución

`GEOMETRY`

Si geom es nulo, entonces se devuelve el valor nulo.

Si geom está vacío, entonces se devuelve el valor null.

Si geom no es `LINESTRING`, entonces se devuelve el valor null.

### Ejemplos

El siguiente SQL devuelve una representación extended well-known text (EWKT) de una `LINESTRING` de cuatro puntos a un objeto `GEOMETRY` y devuelve el punto inicial de la linestring.

```
SELECT ST_AsEWKT(ST_StartPoint(ST_GeomFromText('LINESTRING(0 0,10 0,10 10,5 5,0
 5)',4326)));
```

```
st_asewkt
```

```
-------------
 SRID=4326;POINT(0 0)
```

## ST_Touches

ST_Touches devuelve true si las proyecciones 2D de las dos geometrías de entrada se tocan. Las dos geometrías se tocan si no están vacías, se intersecan y no tienen puntos interiores en común.

### Sintaxis

```
ST_Touches(geom1, geom2)
```

### Argumentos

geom1

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

geom2

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

### Tipo de devolución

`BOOLEAN`

Si geom1 o geom2 son null, entonces se devuelve el valor null.

Si geom1 y geom2 no tienen el mismo valor en el identificador de sistema de referencia espacial (SRID), se devuelve un error.

Si geom1 o geom2 son una colección geométrica, se devuelve un error.

### Ejemplos

El siguiente SQL comprueba si un polígono toca una linestring.

```
SELECT ST_Touches(ST_GeomFromText('POLYGON((0 0,10 0,0 10,0 0))'),
 ST_GeomFromText('LINESTRING(20 10,20 0,10 0)'));
```

```
 st_touches
-------------
```

```
 t
```

## ST_Transform

ST_transform devuelve una nueva geometría con coordenadas que se transforman en un sistema de referencia espacial definido mediante el identificador del sistema de referencia espacial (SRID) de entrada.

### Sintaxis

```
ST_Transform(geom, srid)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

srid

Un valor de tipo de datos `INTEGER` que representa un SRID.

### Tipo de devolución

`GEOMETRY`.

El valor SRID de la geometría devuelta se establece en srid.

Si geom o srid son null, entonces se devuelve el valor null.

Si el valor SRID asociado a la geom de entrada no existe, se devuelve un error.

Si srid no existe, se devuelve un error.

### Ejemplos

El siguiente SQL transforma el SRID de una colección de geometría vacía.

```
SELECT ST_AsEWKT(ST_Transform(ST_GeomFromText('GEOMETRYCOLLECTION EMPTY', 3857),
 4326));
```

```
            st_asewkt
```

```
-----------------------------------
 SRID=4326;GEOMETRYCOLLECTION EMPTY
```

El siguiente SQL transforma el valor SRID de una cadena de líneas.

```
SELECT ST_AsEWKT(ST_Transform(ST_GeomFromText('LINESTRING(110 40, 2 3, -10 80, -7 9,
 -22 -33)', 4326), 26918));
```

```
     st_asewkt
-----------------------------------------------------------------------------------------------
 SRID=26918;LINESTRING(73106.6977300955 15556182.9688576,14347201.5059964
 1545178.32934967,1515090.41262989 9522193.25115316,10491250.83295
 2575457.28410878,5672303.72135968 -5233682.61176205)
```

El siguiente SQL transforma el valor SRID de un polígono.

```
SELECT ST_AsEWKT(ST_Transform(ST_GeomFromText('POLYGON Z ((-10 10 -7, -65 10 -6, -10 64
 -5, -10 10 -7), (-11 11 5, -11 12 6, -12 11 7, -11 11 5))', 6989), 6317));
```

```
                                      st_asewkt
-----------------------------------------------------------------------------------------------
 SRID=6317;POLYGON Z ((6186430.2771091 -1090834.57212608
 1100247.33216237,2654831.67853801 -5693304.90741276 1100247.50581055,2760987.41750022
 -486836.575101877 5709710.44137268,6186430.2771091 -1090834.57212608
 1100247.33216237),(6146675.25029258 -1194792.63532103 1209007.1115113,6125027.87562215
 -1190584.81194058 1317403.77865723,6124888.99555252 -1301885.3455052
 1209007.49312929,6146675.25029258 -1194792.63532103 1209007.1115113))
```

## ST_Union

ST_Union devuelve una geometría que representa la unión de dos geometrías. Es decir, fusiona las geometrías de entrada para producir una geometría resultante sin superposiciones.

Sintaxis

```
ST_Union(geom1, geom2)
```

### Argumentos

geom1

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

geom2

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

### Tipo de devolución

GEOMETRY

El valor del identificador de sistema de referencia espacial (SRID) de la geometría devuelta es el valor SRID de las geometrías de entrada de datos.

Si geom1 o geom2 son null, entonces se devuelve el valor null.

Si geom1 o geom2 están vacíos, se devuelve una geometría vacía.

Si geom1 y geom2 no tienen el mismo valor en el identificador de sistema de referencia espacial (SRID), se devuelve un error.

Si geom1 o geom2 son una recopilación geométrica, cadena de líneas o cadenas de líneas múltiples, se devuelve un error.

Si geom1 o geom2 no es una geometría bidimensional (2D), entonces se devuelve un error.

### Ejemplos

El siguiente SQL devuelve la geometría que no está vacía y representa la unión entre dos geometrías de entrada.

```
SELECT ST_AsEWKT(ST_Union(ST_GeomFromText('POLYGON((0 0,100 100,0 200,0 0))'),
 ST_GeomFromText('POLYGON((0 0,10 0,0 10,0 0))')));
```

```
        st_asewkt
-------------------------
 POLYGON((0 0,0 200,100 100,5 5,10 0,0 0))
```

## ST_Within

ST_Within devuelve true si la proyección 2D de la primera geometría de entrada está dentro de la proyección 2D de la segunda geometría de entrada.

Por ejemplo, la geometría A contiene geometría B si cada punto de A es un punto en B, y sus interiores tienen intersecciones no vacías.

ST_Within(A, B) es equivalente a ST_Contains(B, A).

### Sintaxis

```
ST_Within(geom1, geom2)
```

### Argumentos

geom1

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY. Este valor se compara con geom2 para determinar si está dentro de geom2.

geom2

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

### Tipo de devolución

BOOLEAN

Si geom1 o geom2 son null, entonces se devuelve el valor null.

Si geom1 y geom2 no tienen el mismo identificador de sistema de referencia espacial (SRID), se devuelve un error.

Si geom1 o geom2 son una colección geométrica, se devuelve un error.

### Ejemplos

El siguiente SQL comprueba si el primer polígono está dentro del segundo polígono.

```
SELECT ST_Within(ST_GeomFromText('POLYGON((0 2,1 1,0 -1,0 2))'),
 ST_GeomFromText('POLYGON((-1 3,2 1,0 -3,-1 3))'));
```

```
st_within
-----------
 true
```

## ST_X

ST_X devuelve la primera coordenada de un punto de entrada.

### Sintaxis

```
ST_X(point)
```

### Argumentos

point

Un valor POINT de tipo de datos GEOMETRY.

### Tipo de devolución

Valor DOUBLE PRECISION de la primera coordenada.

Si point es null, entonces se devuelve el valor null.

Si point es un punto vacío, se devuelve el valor null.

Si point no es POINT, se devuelve un error.

### Ejemplos

El siguiente SQL devuelve la primera coordenada de un punto.

```
SELECT ST_X(ST_Point(1,2));
```

```
st_x
-----------
 1.0
```

## ST_XMax

ST_XMax devuelve la máxima primera coordenada de una geometría de entrada de datos.

### Sintaxis

```
ST_XMax(geom)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

### Tipo de devolución

Valor `DOUBLE PRECISION` de la primera coordenada máxima.

Si geom está vacío, entonces se devuelve el valor null.

Si geom es nulo, entonces se devuelve el valor nulo.

### Ejemplos

El siguiente SQL devuelve la primera coordenada más grande de un linestring.

```
SELECT ST_XMax(ST_GeomFromText('LINESTRING(77.29 29.07,77.42 29.26,77.27 29.31,77.29
 29.07)'));
```

```
st_xmax
-----------
 77.42
```

## ST_XMin

ST_XMin devuelve la mínima primera coordenada de una geometría de entrada de datos.

### Sintaxis

```
ST_XMin(geom)
```

### Argumentos

geom

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

### Tipo de devolución

Valor DOUBLE PRECISION de la primera coordenada mínima.

Si geom está vacío, entonces se devuelve el valor null.

Si geom es nulo, entonces se devuelve el valor nulo.

### Ejemplos

El siguiente SQL devuelve la primera coordenada más pequeña de un linestring.

```
SELECT ST_XMin(ST_GeomFromText('LINESTRING(77.29 29.07,77.42 29.26,77.27 29.31,77.29
 29.07)'));
```

```
st_xmin
-----------
 77.27
```

## ST_Y

ST_Y devuelve la segunda coordenada de un punto de entrada.

### Sintaxis

```
ST_Y(point)
```

### Argumentos

point

Un valor POINT de tipo de datos GEOMETRY.

Tipo de devolución

Valor DOUBLE PRECISION de la segunda coordenada.

Si point es null, entonces se devuelve el valor null.

Si point es un punto vacío, se devuelve el valor null.

Si point no es POINT, se devuelve un error.

Ejemplos

El siguiente SQL devuelve la segunda coordenada de un punto.

```
SELECT ST_Y(ST_Point(1,2));
```

```
st_y
-----------
 2.0
```

## ST_YMax

ST_YMax devuelve la máxima segunda coordenada de una geometría de entrada de datos.

Sintaxis

```
ST_YMax(geom)
```

Argumentos

geom

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

Tipo de devolución

Valor DOUBLE PRECISION de la segunda coordenada máxima.

Si geom está vacío, entonces se devuelve el valor null.

Si geom es nulo, entonces se devuelve el valor nulo.

### Ejemplos

El siguiente SQL devuelve la segunda coordenada más grande de un linestring.

```
SELECT ST_YMax(ST_GeomFromText('LINESTRING(77.29 29.07,77.42 29.26,77.27 29.31,77.29
 29.07)'));
```

```
st_ymax
-----------
 29.31
```

## ST_YMin

ST_YMin devuelve la mínima segunda coordenada de una geometría de entrada de datos.

### Sintaxis

```
ST_YMin(geom)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

### Tipo de devolución

Valor `DOUBLE PRECISION` de la segunda coordenada mínima.

Si geom está vacío, entonces se devuelve el valor null.

Si geom es nulo, entonces se devuelve el valor nulo.

### Ejemplos

El siguiente SQL devuelve la segunda coordenada más pequeña de un linestring.

```
SELECT ST_YMin(ST_GeomFromText('LINESTRING(77.29 29.07,77.42 29.26,77.27 29.31,77.29
 29.07)'));
```

```
st_ymin
-----------
 29.07
```

## ST_Z

ST_Z devuelve la coordenada z de un punto de entrada.

### Sintaxis

```
ST_Z(point)
```

### Argumentos

point

Un valor POINT de tipo de datos GEOMETRY.

### Tipo de devolución

Valor DOUBLE PRECISION de la coordenada m.

Si point es null, entonces se devuelve el valor null.

Si point es un punto 2D o 3DM, se devuelve el valor null.

Si point es un punto vacío, se devuelve el valor null.

Si point no es POINT, se devuelve un error.

### Ejemplos

El siguiente SQL devuelve la coordenada z de un punto en una geometría 3DZ.

```
SELECT ST_Z(ST_GeomFromEWKT('POINT Z (1 2 3)'));
```

```
st_z
-----------
 3
```

El siguiente SQL devuelve la coordenada z de un punto en una geometría 4D.

```
SELECT ST_Z(ST_GeomFromEWKT('POINT ZM (1 2 3 4)'));
```

```
st_z
-----------
 3
```

## ST_ZMax

ST_ZMax devuelve la coordenada z máxima de una geometría de entrada.

Sintaxis

```
ST_ZMax(geom)
```

Argumentos

geom

Un valor de tipo de datos GEOMETRY o una expresión que toma el valor de un tipo GEOMETRY.

Tipo de devolución

Valor DOUBLE PRECISION de la primera coordenada z máxima.

Si geom está vacío, entonces se devuelve el valor null.

Si geom es nulo, entonces se devuelve el valor nulo.

Si geom es una geometría 2D o 3DM, se devuelve el valor null.

Ejemplos

El siguiente SQL devuelve la coordenada z más grande de una cadena de línea en una geometría 3DZ.

```
SELECT ST_ZMax(ST_GeomFromEWKT('LINESTRING Z (0 1 2, 3 4 5, 6 7 8)'));
```

```
st_zmax
-----------
  8
```

El siguiente SQL devuelve la coordenada z más grande de una cadena de línea en una geometría 4D.

```
SELECT ST_ZMax(ST_GeomFromEWKT('LINESTRING ZM (0 1 2 3, 4 5 6 7, 8 9 10 11)'));
```

```
st_zmax
-----------
  10
```

## ST_ZMin

ST_ZMin devuelve la coordenada z mínima de una geometría de entrada.

### Sintaxis

```
ST_ZMin(geom)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

### Tipo de devolución

Valor `DOUBLE PRECISION` de la coordenada z mínima.

Si geom está vacío, entonces se devuelve el valor null.

Si geom es nulo, entonces se devuelve el valor nulo.

Si geom es una geometría 2D o 3DM, se devuelve el valor null.

### Ejemplos

El siguiente SQL devuelve la coordenada z más pequeña de una cadena de línea en una geometría 3DZ.

```
SELECT ST_ZMin(ST_GeomFromEWKT('LINESTRING Z (0 1 2, 3 4 5, 6 7 8)'));
```

```
st_zmin
-----------
  2
```

El siguiente SQL devuelve la coordenada z más pequeña de una cadena de línea en una geometría 4D.

```
SELECT ST_ZMin(ST_GeomFromEWKT('LINESTRING ZM (0 1 2 3, 4 5 6 7, 8 9 10 11)'));
```

```
st_zmin
-----------
  2
```

## SupportsBBox

SupportsBBox devuelve true si la geometría de entrada admite la codificación con un cuadro delimitador calculado previamente. Para obtener más información acerca de la compatibilidad con los cuadros delimitadores, consulte Cuadro delimitador.

### Sintaxis

```
SupportsBBox(geom)
```

### Argumentos

geom

Un valor de tipo de datos `GEOMETRY` o una expresión que toma el valor de un tipo `GEOMETRY`.

### Tipo de devolución

`BOOLEAN`

Si geom es nulo, entonces se devuelve el valor nulo.

Ejemplos

El siguiente SQL devuelve true porque la geometría de punto de entrada admite la codificación con un cuadro delimitador.

```
SELECT SupportsBBox(AddBBox(ST_GeomFromText('POLYGON((0 0,1 0,0 1,0 0))')));
```

```
supportsbbox
--------------
t
```

El siguiente SQL devuelve false porque la geometría de punto de entrada no admite la codificación con un cuadro delimitador.

```
SELECT SupportsBBox(DropBBox(ST_GeomFromText('POLYGON((0 0,1 0,0 1,0 0))')));
```

```
supportsbbox
--------------
f
```

## Funciones de cadena

Temas


- || Operador (concatenación)
- Función ASCII
- Función BPCHARCMP
- Función BTRIM
- Función BTTEXT_PATTERN_CMP
- Función CHAR_LENGTH
- Función CHARACTER_LENGTH
- Función CHARINDEX
- Función CHR
- Función COLLATE
- Función CONCAT

- Función CRC32
- Función DIFFERENCE
- Función INITCAP
- Funciones LEFT y RIGHT
- Función LEN
- Función LENGTH
- Función LOWER
- Funciones LPAD y RPAD
- Función LTRIM
- Función OCTETINDEX
- Función OCTET_LENGTH
- Función POSITION
- Función QUOTE_IDENT
- Función QUOTE_LITERAL
- Función REGEXP_COUNT
- Función REGEXP_INSTR
- Función REGEXP_REPLACE
- Función REGEXP_SUBSTR
- Función REPEAT
- Función REPLACE
- Función REPLICATE
- Función REVERSE
- Función RTRIM
- Función SOUNDEX
- Función SPLIT_PART
- Función STRPOS
- Función STRTOL
- Función SUBSTRING
- Función TEXTLEN
- Función TRANSLATE

- Función TRIM
- Función UPPER

Las funciones de cadena procesan y administran cadenas de caracteres o expresiones que tomen el valor de cadenas de caracteres. Cuando el argumento string de estas funciones es un valor literal, debe incluirse entre comillas simples. Entre los tipos de datos compatibles, se incluyen CHAR y VARCHAR.

En la sección siguiente, se proporcionan los nombres de función, la sintaxis y las descripciones para las funciones compatibles. Todos los desplazamientos en cadenas se basan en uno.

Funciones específicas del nodo principal obsoletas

Las siguientes funciones de cadena son obsoletas porque se ejecutan únicamente en el nodo principal. Para obtener más información, consulte Funciones específicas del nodo principal

- GET_BYTE
- SET_BIT
- SET_BYTE
- TO_ASCII

## || Operador (concatenación)

Concatena dos expresiones a ambos extremos del símbolo `||` y devuelve una expresión concatenada.

Similar a Función CONCAT.

> **Note**
>
> Si una o ambas expresiones son nulas, el resultado de la concatenación es `NULL`.

Sintaxis

```
expression1 || expression2
```

### Argumentos

expression1

Una cadena CHAR, una cadena VARCHAR, una expresión binaria o una expresión que se evalúa como uno de estos tipos.

expression2

Una cadena CHAR, una cadena VARCHAR, una expresión binaria o una expresión que se evalúa como uno de estos tipos.

### Tipo de retorno

El tipo de devolución de la cadena es el mismo que el de los argumentos de entrada. Por ejemplo, la concatenación de dos cadenas de tipo VARCHAR devuelve una cadena de tipo VARCHAR.

### Ejemplos

En los siguientes ejemplos se utilizan las tablas USERS y VENUE de la base de datos de muestra de TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para concatenar los campos FIRSTNAME y LASTNAME de la tabla USERS de la base de datos de muestra, utilice el siguiente ejemplo.

```
SELECT (firstname || ' ' || lastname) as fullname
FROM users
ORDER BY 1
LIMIT 10;

+-----------------+
|    fullname     |
+-----------------+
| Aaron Banks     |
| Aaron Booth     |
| Aaron Browning  |
| Aaron Burnett   |
| Aaron Casey     |
| Aaron Cash      |
| Aaron Castro    |
| Aaron Dickerson |
| Aaron Dixon     |
| Aaron Dotson    |
```

```
+-----------------+
```

Para concatenar columnas que puedan llegar a tener valores nulos, use la expresión Funciones NVL y COALESCE. En el siguiente ejemplo se utiliza NVL para devolver `0` siempre que se encuentre `NULL`.

```
SELECT (venuename || ' seats ' || NVL(venueseats, 0)) as seating
FROM venue
WHERE venuestate = 'NV' or venuestate = 'NC'
ORDER BY 1
LIMIT 10;

+-------------------------------------+
|               seating               |
+-------------------------------------+
| Ballys Hotel seats 0                |
| Bank of America Stadium seats 73298 |
| Bellagio Hotel seats 0              |
| Caesars Palace seats 0              |
| Harrahs Hotel seats 0               |
| Hilton Hotel seats 0                |
| Luxor Hotel seats 0                 |
| Mandalay Bay Hotel seats 0          |
| Mirage Hotel seats 0                |
| New York New York seats 0           |
+-------------------------------------+
```

## Función ASCII

La función ASCII devuelve el código ASCII, o el punto de código Unicode, del primer carácter de la cadena especificada. La función devuelve `0` si la cadena está vacía. Si la cadena es nula, se devuelve `NULL`.

### Sintaxis

```
ASCII('string')
```

### Argumento

string

Una cadena `CHAR` o `VARCHAR`.

Tipo de retorno

INTEGER

Ejemplos

Para devolver NULL, use el siguiente ejemplo. La función NULLIF devuelve NULL si los dos argumentos son iguales, así que el argumento de entrada de la función ASCII es NULL. Para obtener más información, consulte Función NULLIF.

```
SELECT ASCII(NULLIF('',''));

+-------+
| ascii |
+-------+
|  NULL |
+-------+
```

Para devolver el código ASCII de 0, utilice el siguiente ejemplo.

```
SELECT ASCII('');

+-------+
| ascii |
+-------+
|     0 |
+-------+
```

Para devolver el código ASCII 97 para la primera letra de la palabra Amazon, use el siguiente ejemplo.

```
SELECT ASCII('amazon');

+-------+
| ascii |
+-------+
|    97 |
+-------+
```

Para devolver el código ASCII 65 para la primera letra de la palabra Amazon, use el siguiente ejemplo.

```
SELECT ASCII('Amazon');

+-------+
| ascii |
+-------+
|    65 |
+-------+
```

## Función BPCHARCMP

Compara el valor de dos cadenas y devuelve un valor entero. Si las cadenas son idénticas, la función devuelve `0`. Si la primera cadena es mayor alfabéticamente, la función devuelve `1`. Si la segunda cadena es mayor, la función devuelve `-1`.

Para caracteres multibyte, la comparación se basa en la codificación del byte.

Sinónimo de Función BTTEXT_PATTERN_CMP.

Sintaxis

```
BPCHARCMP(string1, string2)
```

Argumentos

string1

Una cadena `CHAR` o `VARCHAR`.

string2 (cadena2)

Una cadena `CHAR` o `VARCHAR`.

Tipo de retorno

INTEGER

Ejemplos

En los siguientes ejemplos se utiliza la tabla VENUE de la base de datos de muestra de TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para determinar si el nombre de un usuario es alfabéticamente mayor que el apellido del usuario para las primeras diez entradas en la tabla USERS, use el siguiente ejemplo. Para las entradas en las que la cadena de FIRSTNAME es posterior alfabéticamente a la cadena de LASTNAME, la función devuelve 1. Si LASTNAME es alfabéticamente posterior a FIRSTNAME, la función devuelve -1.

```
SELECT userid, firstname, lastname, BPCHARCMP(firstname, lastname)
FROM users
ORDER BY 1, 2, 3, 4
LIMIT 10;

+--------+-----------+-----------+-----------+
| userid | firstname | lastname  | bpcharcmp |
+--------+-----------+-----------+-----------+
|      1 | Rafael    | Taylor    |        -1 |
|      2 | Vladimir  | Humphrey  |         1 |
|      3 | Lars      | Ratliff   |        -1 |
|      4 | Barry     | Roy       |        -1 |
|      5 | Reagan    | Hodge     |         1 |
|      6 | Victor    | Hernandez |         1 |
|      7 | Tamekah   | Juarez    |         1 |
|      8 | Colton    | Roy       |        -1 |
|      9 | Mufutau   | Watkins   |        -1 |
|     10 | Naida     | Calderon  |         1 |
+--------+-----------+-----------+-----------+
```

Para devolver todas las entradas de la tabla USERS en las que la función devuelve `0`, use el siguiente ejemplo. La función devuelve `0` cuando FIRSTNAME es idéntico a LASTNAME.

```
SELECT userid, firstname, lastname,
BPCHARCMP(firstname, lastname)
FROM users
WHERE BPCHARCMP(firstname, lastname)=0
ORDER BY 1, 2, 3, 4;

+--------+-----------+----------+-----------+
| userid | firstname | lastname | bpcharcmp |
+--------+-----------+----------+-----------+
|     62 | Chase     | Chase    |         0 |
|   4008 | Whitney   | Whitney  |         0 |
|  12516 | Graham    | Graham   |         0 |
|  13570 | Harper    | Harper   |         0 |
```

```
|  16712 | Cooper    | Cooper   |         0 |
|  18359 | Chase     | Chase    |         0 |
|  27530 | Bradley   | Bradley  |         0 |
|  31204 | Harding   | Harding  |         0 |
+--------+-----------+----------+-----------+
```

## Función BTRIM

La función BTRIM recorta una cadena al eliminar espacios o caracteres a la izquierda y a la derecha que coincidan con una cadena específica opcional.

### Sintaxis

```
BTRIM(string [, trim_chars ] )
```

### Argumentos

string

Es la cadena VARCHAR de entrada que se va a recortar.

trim_chars

Es la cadena VARCHAR que contiene los caracteres que deben coincidir.

### Tipo de retorno

La función BTRIM devuelve una cadena VARCHAR.

### Ejemplos

En el siguiente ejemplo, se recortan espacios a la izquierda y a la derecha de la cadena `' abc '`:

```
select '     abc    ' as untrim, btrim('     abc    ') as trim;

untrim    | trim
----------+------
   abc    | abc
```

En el siguiente ejemplo, se eliminan las cadenas `'xyz'` a la izquierda y a la derecha de la cadena `'xyzaxyzbxyzcxyz'`. Las coincidencias a la izquierda y a la derecha de `'xyz'` se eliminan, pero las coincidencias internas dentro de la cadena no se eliminan.

```
select 'xyzaxyzbxyzcxyz' as untrim,
btrim('xyzaxyzbxyzcxyz', 'xyz') as trim;

     untrim      |   trim
-----------------+-----------
 xyzaxyzbxyzcxyz | axyzbxyzc
```

En el siguiente ejemplo, se eliminan las partes a la izquierda y a la derecha de la cadena `'setuphistorycassettes'` que coinciden con cualquiera de los caracteres de la lista trim_chars `'tes'`. Cualquier `t`, `e` o `s` que aparezca antes de cualquier carácter que no esté en la lista trim_chars a la izquierda o a la derecha de la cadena de entrada se eliminará.

```
SELECT btrim('setuphistorycassettes', 'tes');

     btrim
-----------------
 uphistoryca
```

## Función BTTEXT_PATTERN_CMP

Sinónimo para la función BPCHARCMP.

Para obtener más información, consulte Función BPCHARCMP.

## Función CHAR_LENGTH

Sinónimo de la función LEN.

Consulte Función LEN.

## Función CHARACTER_LENGTH

Sinónimo de la función LEN.

Consulte Función LEN.

## Función CHARINDEX

Devuelve la ubicación de la subcadena especificada dentro de una cadena.

Consulte Función POSITION y Función STRPOS para ver funciones similares.

### Sintaxis

```
CHARINDEX( substring, string )
```

### Argumentos

subcadena

Subcadena que se va a buscar dentro de la cadena.

string

La cadena o columna que se buscará.

### Tipo de retorno

INTEGER

La función CHARINDEX devuelve un valor `INTEGER` correspondiente a la posición de la subcadena (basado en uno, no basado en cero). La posición se basa en la cantidad de caracteres, no bytes, por lo que los caracteres multibyte se cuentan como caracteres simples. CHARINDEX devuelve `0` si no se encuentra una subcadena dentro de la cadena.

### Ejemplos

Para devolver la posición de la cadena `fish` dentro de la palabra `dog`, use el siguiente ejemplo.

```
SELECT CHARINDEX('fish', 'dog');

+-----------+
| charindex |
+-----------+
|         0 |
+-----------+
```

Para devolver la posición de la cadena `fish` dentro de la palabra `dogfish`, use el siguiente ejemplo.

```
SELECT CHARINDEX('fish', 'dogfish');
```

```
+-----------+
| charindex |
+-----------+
|         4 |
+-----------+
```

En el siguiente ejemplo, se utiliza la tabla SALES de la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para devolver la cantidad de transacciones de venta distintas con un parámetro de comisión que supere los 999 de la tabla SALES, use el siguiente ejemplo. Este comando cuenta las comisiones superiores a 999 comprobando si el decimal está a más de 4 lugares del principio del valor de la comisión.

```
SELECT DISTINCT CHARINDEX('.', commission), COUNT (CHARINDEX('.', commission))
FROM sales
WHERE CHARINDEX('.', commission) > 4
GROUP BY CHARINDEX('.', commission)
ORDER BY 1,2;

+-----------+-------+
| charindex | count |
+-----------+-------+
|         5 |   629 |
+-----------+-------+
```

## Función CHR

La función CHR devuelve el carácter que coincida con el valor de punto de código ASCII especificado por el parámetro de entrada.

### Sintaxis

```
CHR(number)
```

### Argumento

número

El parámetro de entrada es un valor `INTEGER` que representa un valor de punto de código ASCII.

Tipo de retorno

CHAR

La función CHR devuelve una cadena CHAR si el carácter ASCII coincide con el valor de entrada. Si el número de entrada no tiene una coincidencia ASCII, la función devuelve un valor NULL.

Ejemplos

Para devolver el carácter que corresponde al punto 0 del código ASCII, utilice el siguiente ejemplo. Tenga en cuenta que la función CHR devuelve NULL para la entrada 0.

```
SELECT CHR(0);

+-----+
| chr |
+-----+
|     |
+-----+
```

Para devolver el carácter que corresponde al punto 65 del código ASCII, utilice el siguiente ejemplo.

```
SELECT CHR(65);

+-----+
| chr |
+-----+
| A   |
+-----+
```

Para devolver distintos nombres de eventos que comienzan con una A mayúscula (punto de código ASCII 65). use el siguiente ejemplo. En el siguiente ejemplo se utiliza la tabla EVENT de la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

```
SELECT DISTINCT eventname FROM event
WHERE SUBSTRING(eventname, 1, 1)=CHR(65) LIMIT 5;

+-----------------------+
|       eventname       |
+-----------------------+
| A Catered Affair      |
```

```
| As You Like It        |
| A Man For All Seasons |
| Alan Jackson          |
| Armando Manzanero     |
+-----------------------+
```

## Función COLLATE

La función COLLATE anula la intercalación de una expresión o una columna de cadenas.

Para obtener información sobre cómo crear tablas mediante la intercalación de bases de datos, consulte CREATE TABLE.

Para obtener información sobre cómo crear bases de datos mediante la intercalación de bases de datos, consulte CREATE DATABASE.

### Sintaxis

```
COLLATE( string, 'case_sensitive' | 'case_insensitive');
```

### Argumentos

string

Una expresión o una columna de cadenas que desea invalidar.

“case_sensitive” | “case_insensitive”

Una constante de cadena de un nombre de intercalación. Amazon Redshift solo admite case_sensitive o case_insensitive.

### Tipo de retorno

La función COLLATE devuelve `VARCHAR` o `CHAR` de acuerdo con el primer tipo de expresión de entrada. Esta función solo cambia la intercalación del primer argumento de entrada y no modifica su valor de salida.

### Ejemplos

Para crear la tabla T y definir col1 en la tabla T como `case_sensitive`, utilice el ejemplo siguiente.

```
CREATE TABLE T ( col1 Varchar(20) COLLATE case_sensitive );
```

```
INSERT INTO T VALUES ('john'),('JOHN');
```

Al ejecutar la primera consulta, Amazon Redshift solo devuelve `john`. Después de que la función COLLATE se ejecuta en col1, la intercalación se vuelve `case_insensitive`. La segunda consulta devuelve `john` y `JOHN`.

```
SELECT * FROM T WHERE col1 = 'john';

+------+
| col1 |
+------+
| john |
+------+

SELECT * FROM T WHERE COLLATE(col1, 'case_insensitive') = 'john';

+------+
| col1 |
+------+
| john |
| JOHN |
+------+
```

Para crear la tabla A y definir col1 en la tabla A como `case_insensitive`, utilice el ejemplo siguiente.

```
CREATE TABLE A ( col1 Varchar(20) COLLATE case_insensitive );

INSERT INTO A VALUES ('john'),('JOHN');
```

Al ejecutar la primera consulta, Amazon Redshift devuelve `john` y `JOHN`. Después de que la función COLLATE se ejecuta en col1, la intercalación se vuelve `case_sensitive`. La segunda consulta solo devuelve `john`.

```
SELECT * FROM A WHERE col1 = 'john';

+------+
| col1 |
+------+
| john |
| JOHN |
```

```
+------+

SELECT * FROM A WHERE COLLATE(col1, 'case_sensitive') = 'john';

+------+
| col1 |
+------+
| john |
+------+
```

## Función CONCAT

La función CONCAT concatena dos expresiones y devuelve la expresión resultante. Para concatenar más de dos expresiones, utilice las funciones CONCAT anidadas. El operador de concatenación (| |) entre dos expresiones produce los mismos resultados que la función CONCAT.

### Sintaxis

```
CONCAT ( expression1, expression2 )
```

### Argumentos

expression1, expression2

Ambos argumentos pueden consistir en una cadena de caracteres de longitud fija, una cadena de caracteres de longitud variable, una expresión binaria o una expresión que tiene como valor una de estas entradas de datos.

### Tipo de retorno

CONCAT devuelve una expresión. El tipo de datos de la expresión es igual al de los argumentos de entrada.

Si las expresiones de entrada son de diferentes tipos, Amazon Redshift intenta realizar una conversión de tipo de manera implícita en una de las expresiones. Si no se pueden convertir los valores, se devuelve un error.

### Notas de uso

- Para la función CONCAT y el operador de concatenación, si una o ambas expresiones son nulas, el resultado de la concatenación también lo será.

## Ejemplos

En el siguiente ejemplo, se concatenan dos literales de caracteres:

```
SELECT CONCAT('December 25, ', '2008');

concat
-------------------
December 25, 2008
(1 row)
```

La siguiente consulta, utilizando el operador || en lugar de CONCAT, produce el mismo resultado:

```
SELECT 'December 25, '||'2008';

?column?
-------------------
December 25, 2008
(1 row)
```

El siguiente ejemplo utiliza una función CONCAT anidada dentro de otra función CONCAT para concatenar tres cadenas de caracteres:

```
SELECT CONCAT('Thursday, ', CONCAT('December 25, ', '2008'));

concat
----------------------------
Thursday, December 25, 2008
(1 row)
```

Para concatenar columnas que puedan contener NULL, utilice la función Funciones NVL y COALESCE, que devuelve un valor determinado cuando encuentra un valor NULL. En el siguiente ejemplo, se usa NVL para devolver un 0 siempre que se encuentre un NULL.

```
SELECT CONCAT(venuename, CONCAT(' seats ', NVL(venueseats, 0))) AS seating
FROM venue WHERE venuestate = 'NV' OR venuestate = 'NC'
ORDER BY 1
LIMIT 5;

seating
-----------------------------------
```

```
Ballys Hotel seats 0
Bank of America Stadium seats 73298
Bellagio Hotel seats 0
Caesars Palace seats 0
Harrahs Hotel seats 0
(5 rows)
```

En la siguiente consulta, se concatenan valores CITY y STATE de la tabla VENUE:

```
SELECT CONCAT(venuecity, venuestate)
FROM venue
WHERE venueseats > 75000
ORDER BY venueseats;

concat
-------------------
DenverCO
Kansas CityMO
East RutherfordNJ
LandoverMD
(4 rows)
```

La siguiente consulta utiliza funciones CONCAT anidadas. La consulta concatena los valores CITY y STATE de la tabla VENUE pero delimita la cadena resultado con una coma y un espacio:

```
SELECT CONCAT(CONCAT(venuecity,', '),venuestate)
FROM venue
WHERE venueseats > 75000
ORDER BY venueseats;

concat
---------------------
Denver, CO
Kansas City, MO
East Rutherford, NJ
Landover, MD
(4 rows)
```

En el siguiente ejemplo, se concatenan dos expresiones binarias. Donde `abc` es un valor binario (con una representación hexadecimal de 616263) y `def` es un valor binario (con representación hexadecimal de 646566). El resultado se muestra automáticamente como representación hexadecimal del valor binario.

```
SELECT CONCAT('abc'::VARBYTE, 'def'::VARBYTE);

concat
-------------------
616263646566
```

## Función CRC32

CRC32 es una función que se utiliza para la detección de errores. La función utiliza un algoritmo CRC32 para detectar cambios entre datos de origen y de destino. La función CRC32 convierte una cadena de longitud variable en una cadena de 8 caracteres que es una representación textual del valor hexadecimal de una secuencia binaria de 32 bits. Para detectar cambios entre los datos de origen y de destino, utilice la función CRC32 en los datos de origen y almacene la salida. A continuación, utilice la función CRC32 en los datos de destino y compare esa salida con la salida de los datos de origen. Las salidas serán las mismas si los datos no se modificaron y las salidas serán diferentes si se modificaron los datos.

### Sintaxis

```
CRC32(string)
```

### Argumentos

string

Una cadena CHAR, una cadena VARCHAR o una expresión que se evalúe de forma implícita como un tipo CHAR o VARCHAR.

### Tipo de retorno

La función CRC32 devuelve una cadena de 8 caracteres que es una representación textual del valor hexadecimal de una secuencia binaria de 32 bits. La función CRC32 de Amazon Redshift se basa en el polinomio CRC-32C.

### Ejemplos

Para mostrar el valor de 8 bits de la cadena `Amazon Redshift`.

```
SELECT CRC32('Amazon Redshift');
```

```
+----------+
|  crc32   |
+----------+
| f2726906 |
+----------+
```

## Función DIFFERENCE

La función DIFFERENCE compara los códigos de American Soundex de dos cadenas. La función devuelve un `INTEGER` para indicar el número de caracteres coincidentes entre los códigos de Soundex.

Un código de Soundex es una cadena de cuatro caracteres. Un código de Soundex representa cómo suena una palabra en lugar de cómo se deletrea. Por ejemplo, `Smith` y `Smyth` tienen el mismo código de Soundex.

Sintaxis

```
DIFFERENCE(string1, string2)
```

Argumentos

string1

Una cadena `CHAR`, una cadena `VARCHAR` o una expresión que se evalúe de forma implícita como un tipo `CHAR` o `VARCHAR`.

string2 (cadena2)

Una cadena `CHAR`, una cadena `VARCHAR` o una expresión que se evalúe de forma implícita como un tipo `CHAR` o `VARCHAR`.

Tipo de retorno

INTEGER

La función DIFFERENCE devuelve un valor `INTEGER` de 0 a 4 que cuenta el número de caracteres coincidentes en los códigos American Soundex de las dos cadenas. Un código de Soundex tiene 4 caracteres, por lo que la función DIFFERENCE devuelve `4` cuando los

4 caracteres de los valores de código de American Soundex de las cadenas son iguales. DIFFERENCE devuelve `0` si una de las dos cadenas está vacía. La función devuelve 1 si ninguna cadena contiene caracteres válidos. La función DIFFERENCE solo convierte caracteres ASCII alfabéticos en minúsculas o mayúsculas en inglés, incluidas las letras a-z y A-Z. DIFFERENCE omite otros caracteres.

### Ejemplos

Para comparar los valores de Soundex de las cadenas `%` y `@`, utilice el ejemplo siguiente. La función devuelve 1 porque ninguna cadena contiene caracteres válidos.

```
SELECT DIFFERENCE('%', '@');

+------------+
| difference |
+------------+
|          1 |
+------------+
```

Para comparar los valores de Soundex de `Amazon` y una cadena vacía, utilice el ejemplo siguiente. La función devuelve `0` porque una de las dos cadenas está vacía.

```
SELECT DIFFERENCE('Amazon', '');

+------------+
| difference |
+------------+
|          0 |
+------------+
```

Para comparar los valores de Soundex de las cadenas `Amazon` y `Ama`, utilice el ejemplo siguiente. La función devuelve 2 porque 2 caracteres de los valores de Soundex de las cadenas son iguales.

```
SELECT DIFFERENCE('Amazon', 'Ama');

+------------+
| difference |
+------------+
|          2 |
+------------+
```

Para comparar los valores de Soundex de las cadenas `Amazon` y `+-*/%Amazon`, utilice el ejemplo siguiente. La función devuelve `4` porque 4 caracteres de los valores de Soundex de las cadenas son iguales. Observe que la función ignora los caracteres no válidos `+-*/%` en la segunda cadena.

```
SELECT DIFFERENCE('Amazon', '+-*/%Amazon');

+------------+
| difference |
+------------+
|          4 |
+------------+
```

Para comparar los valores de Soundex de las cadenas `AC/DC` y `Ay See Dee See`, utilice el ejemplo siguiente. La función devuelve `4` porque 4 caracteres de los valores de Soundex de las cadenas son iguales.

```
SELECT DIFFERENCE('AC/DC', 'Ay See Dee See');

+------------+
| difference |
+------------+
|          4 |
+------------+
```

## Función INITCAP

Pone en mayúsculas la primera letra de cada palabra en una cadena especificada. INITCAP admite caracteres multibyte UTF-8 de hasta un máximo de cuatro bytes por carácter.

Sintaxis

```
INITCAP(string)
```

Argumento

string

Una cadena `CHAR`, una cadena `VARCHAR` o una expresión que se evalúe de forma implícita como un tipo `CHAR` o `VARCHAR`.

Tipo de retorno

VARCHAR

Notas de uso

La función INITCAP hace que la primera letra de cada palabra en una cadena esté en mayúsculas y todas las letras subsiguientes (o hacia la izquierda) estén en minúsculas. Por lo tanto, es importante comprender qué caracteres (que no sean espacios) funcionan como separadores de palabras. Un carácter separador de palabra es cualquier carácter no alfanumérico, incluidos los signos de puntuación, los símbolos y los caracteres de control. Todos los siguientes caracteres son separadores de palabras:

```
! " # $ % & ' ( ) * + , - . / : ; < = > ? @ [ \ ] ^ _ ` { | } ~
```

Tabulaciones, caracteres en línea nueva, fuentes de forma, fuentes de línea y saltos de línea también son separadores de palabra.

Ejemplos

En los siguientes ejemplos se utilizan datos de las tablas CATEGORY y USERS en la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para escribir en mayúsculas las iniciales de cada palabra en la columna CATDESC, use el siguiente ejemplo.

```
SELECT catid, catdesc, INITCAP(catdesc)
FROM category
ORDER BY 1, 2, 3;

+-------+----------------------------------------------
+-----------------------------------------+
| catid |                  catdesc                  |                  initcap
          |
+-------+----------------------------------------------
+-----------------------------------------+
|     1 | Major League Baseball                     | Major League Baseball
          |
|     2 | National Hockey League                    | National Hockey League
          |
|     3 | National Football League                  | National Football League
          |
```

```
|     4 | National Basketball Association           | National Basketball Association
          |
|     5 | Major League Soccer                       | Major League Soccer
          |
|     6 | Musical theatre                           | Musical Theatre
          |
|     7 | All non-musical theatre                   | All Non-Musical Theatre
          |
|     8 | All opera and light opera                 | All Opera And Light Opera
          |
|     9 | All rock and pop music concerts           | All Rock And Pop Music Concerts
          |
|    10 | All jazz singers and bands                | All Jazz Singers And Bands
          |
|    11 | All symphony, concerto, and choir concerts | All Symphony, Concerto, And
 Choir Concerts |
+-------+-------------------------------------------
+----------------------------------------+
```

Para mostrar que la función INITCAP no preserva caracteres en mayúsculas donde no comienzan palabras, use el siguiente ejemplo. Por ejemplo, la cadena `MLB` se convierte en `Mlb`.

```
SELECT INITCAP(catname)
FROM category
ORDER BY catname;

+-----------+
|  initcap  |
+-----------+
| Classical |
| Jazz      |
| Mlb       |
| Mls       |
| Musicals  |
| Nba       |
| Nfl       |
| Nhl       |
| Opera     |
| Plays     |
| Pop       |
+-----------+
```

Para mostrar que los caracteres no alfanuméricos distintos de los espacios funcionan como separadores de palabras, utilice el siguiente ejemplo. Se escribirán en mayúscula varias letras de cada cadena.

```
SELECT email, INITCAP(email)
FROM users
ORDER BY userid DESC LIMIT 5;

+-----------------------------------+-----------------------------------+
|               email               |              initcap              |
+-----------------------------------+-----------------------------------+
| urna.Ut@egetdictumplacerat.edu    | Urna.Ut@Egetdictumplacerat.Edu    |
| nibh.enim@egestas.ca              | Nibh.Enim@Egestas.Ca              |
| in@Donecat.ca                     | In@Donecat.Ca                     |
| sodales@blanditviverraDonec.ca    | Sodales@Blanditviverradonec.Ca    |
| sociis.natoque.penatibus@vitae.org | Sociis.Natoque.Penatibus@Vitae.Org |
+-----------------------------------+-----------------------------------+
```

## Funciones LEFT y RIGHT

Estas funciones devuelven la cantidad especificada de caracteres más a la izquierda o más a la derecha de una cadena de caracteres.

La cantidad se basa en la cantidad de caracteres, no bytes, por lo que los caracteres multibyte se cuentan como caracteres simples.

Sintaxis

```
LEFT( string,  integer )

RIGHT( string,  integer )
```

Argumentos

string

Una cadena CHAR, una cadena VARCHAR o cualquier expresión que se evalúe como una cadena CHAR o VARCHAR.

integer

Un número entero.

Tipo de retorno

VARCHAR

Ejemplos

En el siguiente ejemplo se utilizan datos de la tabla EVENT en la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para devolver los 5 caracteres más a la izquierda y los 5 más a la derecha de nombres de eventos que tienen ID de eventos entre 1000 y 1005, use el siguiente ejemplo.

```
SELECT eventid, eventname,
LEFT(eventname,5) AS left_5,
RIGHT(eventname,5) AS right_5
FROM event
WHERE eventid BETWEEN 1000 AND 1005
ORDER BY 1;

+---------+----------------+--------+---------+
| eventid |   eventname    | left_5 | right_5 |
+---------+----------------+--------+---------+
|    1000 | Gypsy          | Gypsy  | Gypsy   |
|    1001 | Chicago        | Chica  | icago   |
|    1002 | The King and I | The K  | and I   |
|    1003 | Pal Joey       | Pal J  | Joey    |
|    1004 | Grease         | Greas  | rease   |
|    1005 | Chicago        | Chica  | icago   |
+---------+----------------+--------+---------+
```

## Función LEN

Devuelve la longitud de la cadena especificada como el número de caracteres.

Sintaxis

LEN es sinónimo de Función LENGTH, Función CHAR_LENGTH, Función CHARACTER_LENGTH y Función TEXTLEN.

```
LEN(expression)
```

Argumento

expresión

Una cadena CHAR, una cadena VARCHAR, una expresión VARBYTE o una expresión que se evalúe de forma implícita como un tipo CHAR, VARCHAR o VARBYTE.

Tipo de retorno

INTEGER

La función LEN devuelve un valor entero que indica la cantidad de caracteres en la cadena de entrada.

Si la cadena de entrada es una cadena de caracteres, la función LEN devuelve una cantidad real de caracteres en cadenas multibyte y no la cantidad de bytes. Por ejemplo, una columna VARCHAR(12) necesita almacenar tres caracteres chinos de cuatro bytes. La función LEN devolverá 3 para esa misma cadena. Para obtener la longitud de la cadena en bytes, use la función OCTET_LENGTH.

Notas de uso

Si la expresión es una cadena CHAR, los espacios finales no se cuentan.

Si la expresión es una cadena VARCHAR, los espacios finales se cuentan.

Ejemplos

Para devolver el número de bytes y el número de caracteres de la cadena `français`, use el siguiente ejemplo.

```
SELECT OCTET_LENGTH('français'),
LEN('français');

+--------------+-----+
| octet_length | len |
+--------------+-----+
|            9 |   8 |
+--------------+-----+
```

Para que se devuelva el número de bytes y el número de caracteres de la cadena `français` sin utilizar la función OCTET_LENGTH, utilice el siguiente ejemplo. Para obtener más información, consulte Función CAST.

```
SELECT LEN(CAST('français' AS VARBYTE)) as bytes, LEN('français');

+-------+-----+
| bytes | len |
+-------+-----+
|     9 |   8 |
+-------+-----+
```

Para devolver el número de caracteres de las cadenas `cat` sin espacios finales, `cat` con tres espacios finales, `cat` con tres espacios finales configurados como CHAR de 6 de longitud y `cat` con tres espacios finales moldeados a modo de VARCHAR de longitud 6, utilice el siguiente ejemplo. Observe que la función no cuenta los espacios finales para las cadenas CHAR, pero sí cuenta los espacios finales para las cadenas VARCHAR.

```
SELECT LEN('cat'), LEN('cat   '), LEN(CAST('cat   ' AS CHAR(6))) AS len_char,
 LEN(CAST('cat   ' AS VARCHAR(6))) AS len_varchar;

+-----+-----+----------+-------------+
| len | len | len_char | len_varchar |
+-----+-----+----------+-------------+
|   3 |   6 |        3 |           6 |
+-----+-----+----------+-------------+
```

En el siguiente ejemplo se utilizan datos de la tabla VENUE en la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para devolver los 10 nombres de lugares más largos en la tabla VENUE, utilice el siguiente ejemplo.

```
SELECT venuename, LEN(venuename)
FROM venue
ORDER BY 2 DESC, 1
LIMIT 10;

+----------------------------------------+-----+
|               venuename                | len |
+----------------------------------------+-----+
| Saratoga Springs Performing Arts Center |  39 |
```

```
| Lincoln Center for the Performing Arts  |  38 |
| Nassau Veterans Memorial Coliseum       |  33 |
| Jacksonville Municipal Stadium          |  30 |
| Rangers BallPark in Arlington           |  29 |
| University of Phoenix Stadium           |  29 |
| Circle in the Square Theatre            |  28 |
| Hubert H. Humphrey Metrodome            |  28 |
| Oriole Park at Camden Yards             |  27 |
| Dick's Sporting Goods Park              |  26 |
+-----------------------------------------+-----+
```

## Función LENGTH

Sinónimo de la función LEN.

Consulte Función LEN.

## Función LOWER

Convierte una cadena de caracteres a minúsculas. LOWER admite caracteres multibyte UTF-8 de hasta un máximo de cuatro bytes por carácter.

Sintaxis

```
LOWER(string)
```

Argumento

string

Una cadena o expresión VARCHAR que se evalúa como el tipo VARCHAR.

Tipo de retorno

cadena

La función LOWER devuelve una cadena que presenta el mismo tipo de datos que la cadena de entrada. Por ejemplo, si la entrada es una cadena CHAR, la función devolverá una cadena CHAR.

Ejemplos

En el siguiente ejemplo se utilizan datos de la tabla CATEGORY en la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para convertir las cadenas VARCHAR de la columna CATNAME en minúsculas, utilice el siguiente ejemplo.

```
SELECT catname, LOWER(catname) FROM category ORDER BY 1,2;

+-----------+-----------+
|  catname  |   lower   |
+-----------+-----------+
| Classical | classical |
| Jazz      | jazz      |
| MLB       | mlb       |
| MLS       | mls       |
| Musicals  | musicals  |
| NBA       | nba       |
| NFL       | nfl       |
| NHL       | nhl       |
| Opera     | opera     |
| Plays     | plays     |
| Pop       | pop       |
+-----------+-----------+
```

## Funciones LPAD y RPAD

Estas funciones anteponen o anexan caracteres a una cadena, según una longitud especificada.

Sintaxis

```
LPAD(string1, length, [ string2 ])
```

```
RPAD(string1, length, [ string2 ])
```

Argumentos

string1

Una cadena CHAR, una cadena VARCHAR o una expresión que se evalúe de forma implícita como un tipo CHAR o VARCHAR.

longitud

Un valor entero que define la longitud del resultado de la función. La longitud de una cadena se basa en la cantidad de caracteres, no bytes, por lo que los caracteres multibyte se cuentan como caracteres simples. Si string1 (cadena1) tiene una longitud mayor que la especificada, se trunca (a la derecha). Si el valor de longitud es cero o un número negativo, el resultado de la función es una cadena vacía.

string2 (cadena2)

(Opcional) Uno o varios caracteres que se anteponen o anexan a string1. Si este argumento no se especifica, se utilizan espacios.

Tipo de retorno

VARCHAR

Ejemplos

En los siguientes ejemplos se utilizan datos de la tabla EVENT en la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para truncar un conjunto especificado de nombres de eventos a 20 caracteres y anteponer espacios a los nombres más cortos, use el siguiente ejemplo.

```
SELECT LPAD(eventname, 20) FROM event
WHERE eventid BETWEEN 1 AND 5 ORDER BY 1;

+----------------------+
|         lpad         |
+----------------------+
|               Salome |
|        Il Trovatore |
|       Boris Godunov |
|     Gotterdammerung |
|La Cenerentola (Cind |
+----------------------+
```

Para truncar el mismo conjunto de nombres de eventos a 20 caracteres, pero anexar los nombres más cortos con `0123456789`, use el siguiente ejemplo.

```
SELECT RPAD(eventname, 20,'0123456789') FROM event
```

```
WHERE eventid BETWEEN 1 AND 5 ORDER BY 1;

+----------------------+
|         rpad         |
+----------------------+
| Boris Godunov0123456 |
| Gotterdammerung01234 |
| Il Trovatore01234567 |
| La Cenerentola (Cind |
| Salome01234567890123 |
+----------------------+
```

## Función LTRIM

Recorta los caracteres desde el principio de una cadena. Elimina la cadena más larga que contiene solo caracteres de la lista de caracteres de recorte. El recorte finaliza cuando no aparece ningún carácter de recorte en la cadena de entrada.

### Sintaxis

```
LTRIM( string [, trim_chars] )
```

### Argumentos

string

Una columna de cadena, una expresión o un literal de cadena que se va a recortar.

trim_chars

Una columna de cadena, expresión o literal de cadena que representa los caracteres que se van a recortar desde el principio de la cadena. Si no se especifica, se utiliza un espacio como carácter de recorte.

### Tipo de retorno

La función LTRIM devuelve una cadena de caracteres con el mismo tipo de datos que la cadena de entrada (CHAR o VARCHAR).

Ejemplos

En el siguiente ejemplo, se recorta el año de la columna `listime`. Los caracteres de recorte del literal de cadena `'2008-'` indican los caracteres que se recortarán desde la izquierda. Si utiliza los caracteres de recorte `'028-'`, obtendrá el mismo resultado.

```
select listid, listtime, ltrim(listtime, '2008-')
from listing
order by 1, 2, 3
limit 10;

listid |      listtime       |     ltrim
-------+---------------------+---------------
     1 | 2008-01-24 06:43:29 | 1-24 06:43:29
     2 | 2008-03-05 12:25:29 | 3-05 12:25:29
     3 | 2008-11-01 07:35:33 | 11-01 07:35:33
     4 | 2008-05-24 01:18:37 | 5-24 01:18:37
     5 | 2008-05-17 02:29:11 | 5-17 02:29:11
     6 | 2008-08-15 02:08:13 | 15 02:08:13
     7 | 2008-11-15 09:38:15 | 11-15 09:38:15
     8 | 2008-11-09 05:07:30 | 11-09 05:07:30
     9 | 2008-09-09 08:03:36 | 9-09 08:03:36
    10 | 2008-06-17 09:44:54 | 6-17 09:44:54
```

LTRIM elimina cualquiera de los caracteres de trim_chars cuando aparecen al principio de la cadena. En el siguiente ejemplo, se recortan los caracteres «C», «D» y «G» cuando aparecen al principio de VENUENAME, que es una columna VARCHAR.

```
select venueid, venuename, ltrim(venuename, 'CDG')
from venue
where venuename like '%Park'
order by 2
limit 7;

venueid | venuename                  | btrim
--------+----------------------------+-------------------------
    121 | ATT Park                   | ATT Park
    109 | Citizens Bank Park         | itizens Bank Park
    102 | Comerica Park              | omerica Park
      9 | Dick's Sporting Goods Park | ick's Sporting Goods Park
     97 | Fenway Park                | Fenway Park
    112 | Great American Ball Park   | reat American Ball Park
```

```
    114 | Miller Park               | Miller Park
```

En el siguiente ejemplo. se utiliza el carácter de recorte 2 que se recupera de la columna `venueid`.

```
select ltrim('2008-01-24 06:43:29', venueid)
from venue where venueid=2;

ltrim
------------------
008-01-24 06:43:29
```

En el siguiente ejemplo, no se recorta ningún carácter porque se encuentra un 2 antes del carácter de recorte `'0'`.

```
select ltrim('2008-01-24 06:43:29', '0');

ltrim
-------------------
2008-01-24 06:43:29
```

En el siguiente ejemplo, se utiliza el carácter de recorte de espacio predeterminado y se recortan los dos espacios desde el principio de la cadena.

```
select ltrim('  2008-01-24 06:43:29');

ltrim
-------------------
2008-01-24 06:43:29
```

## Función OCTETINDEX

La función OCTETINDEX devuelve la ubicación de una subcadena dentro de una cadena como un número de bytes.

Sintaxis

```
OCTETINDEX(substring, string)
```

### Argumentos

subcadena

Una cadena CHAR, una cadena VARCHAR o una expresión que se evalúe de forma implícita como un tipo CHAR o VARCHAR.

string

Una cadena CHAR, una cadena VARCHAR o una expresión que se evalúe de forma implícita como un tipo CHAR o VARCHAR.

### Tipo de retorno

INTEGER

La función OCTETINDEX devuelve un valor INTEGER que corresponde a la posición de la subcadena dentro de la cadena como un número de bytes, donde el primer carácter de la cadena se cuenta como 1. Si la cadena no contiene caracteres multibyte, el resultado es igual al resultado de la función CHARINDEX. Si la cadena no contiene la subcadena, la función devuelve 0. Si la subcadena está vacía, la función devuelve 1.

### Ejemplos

Para devolver la posición de la subcadena q en la cadena Amazon Redshift, utilice el ejemplo siguiente. Este ejemplo devuelve 0 porque la subcadena no está en la cadena.

```
SELECT OCTETINDEX('q', 'Amazon Redshift');

+------------+
| octetindex |
+------------+
|          0 |
+------------+
```

Para devolver la posición de una subcadena vacía en la cadena Amazon Redshift, utilice el ejemplo siguiente. Este ejemplo devuelve 1 porque la subcadena está vacía.

```
SELECT OCTETINDEX('', 'Amazon Redshift');
```

```
+------------+
| octetindex |
+------------+
|          1 |
+------------+
```

Para devolver la posición de la subcadena `Redshift` en la cadena `Amazon Redshift`, utilice el ejemplo siguiente. En este ejemplo se devuelve 8 porque la subcadena comienza en el octavo byte de la cadena.

```
SELECT OCTETINDEX('Redshift', 'Amazon Redshift');

+------------+
| octetindex |
+------------+
|          8 |
+------------+
```

Para devolver la posición de la subcadena `Redshift` en la cadena `Amazon Redshift`, utilice el ejemplo siguiente. Este ejemplo devuelve 21 porque los seis primeros caracteres de la cadena son caracteres de dos bytes.

```
SELECT OCTETINDEX('Redshift', 'Άμαζον Amazon Redshift');

+------------+
| octetindex |
+------------+
|         21 |
+------------+
```

## Función OCTET_LENGTH

Devuelve la longitud de la cadena especificada como el número de bytes.

Sintaxis

```
OCTET_LENGTH(expression)
```

Argumento

expresión

Una cadena CHAR, una cadena VARCHAR, una expresión VARBYTE o una expresión que se evalúe de forma implícita como un tipo CHAR, VARCHAR o VARBYTE.

Tipo de retorno

INTEGER

La función OCTET_LENGTH devuelve un valor entero que indica la cantidad de bytes en la cadena de entrada.

Si la cadena de entrada es una cadena de caracteres, la función LEN devuelve una cantidad real de caracteres en cadenas multibyte y no la cantidad de bytes. Por ejemplo, una columna VARCHAR(12) necesita almacenar tres caracteres chinos de cuatro bytes. La función OCTET_LENGTH devolverá 12 para esa cadena, y la función LEN devolverá 3 para esa misma cadena.

Notas de uso

Si expresión es una CHAR, la función devuelve la longitud de la cadena CHAR. Por ejemplo, la salida de una entrada CHAR(6) es CHAR(6).

Si la expresión es una cadena VARCHAR, los espacios finales se cuentan.

Ejemplos

Para devolver el número de bytes cuando la cadena `francais` con tres espacios finales se convierte en CHAR y un tipo de VARCHAR, utilice el siguiente ejemplo. Para obtener más información, consulte Función CAST.

```
SELECT OCTET_LENGTH(CAST('francais   ' AS CHAR(15))) AS octet_length_char,
 OCTET_LENGTH(CAST('francais   ' AS VARCHAR(15))) AS octet_length_varchar;

+------------------+----------------------+
| octet_length_char | octet_length_varchar |
+------------------+----------------------+
|               15 |                   11 |
```

```
+------------------+---------------------+
```

Para devolver el número de bytes y el número de caracteres de la cadena `français`, use el siguiente ejemplo.

```
SELECT OCTET_LENGTH('français'), LEN('français');

+--------------+-----+
| octet_length | len |
+--------------+-----+
|            9 |   8 |
+--------------+-----+
```

Para devolver el número de bytes cuando la cadena `français` se convierte en VARBYTE, utilice el siguiente ejemplo.

```
SELECT OCTET_LENGTH(CAST('français' AS VARBYTE));

+--------------+
| octet_length |
+--------------+
|            9 |
+--------------+
```

## Función POSITION

Devuelve la ubicación de la subcadena especificada dentro de una cadena.

Consulte Función CHARINDEX y Función STRPOS para ver funciones similares.

Sintaxis

```
POSITION(substring IN string )
```

Argumentos

subcadena

Subcadena que se va a buscar dentro de la cadena.

string

La cadena o columna que se buscará.

Tipo de retorno

La función POSITION devuelve un valor `INTEGER` correspondiente a la posición de la subcadena (basado en uno, no basado en cero). La posición se basa en la cantidad de caracteres, no bytes, por lo que los caracteres multibyte se cuentan como caracteres simples. POSITION devuelve `0` si no se encuentra subcadena dentro de la cadena.

Ejemplos

Para devolver la posición de la cadena `fish` dentro de la palabra `dog`, use el siguiente ejemplo.

```
SELECT POSITION('fish' IN 'dog');

+-----------+
|  position |
+-----------+
|         0 |
+-----------+
```

Para devolver la posición de la cadena `fish` dentro de la palabra `dogfish`, use el siguiente ejemplo.

```
SELECT POSITION('fish' IN 'dogfish');

+-----------+
|  position |
+-----------+
|         4 |
+-----------+
```

En el siguiente ejemplo, se utiliza la tabla SALES de la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para devolver la cantidad de transacciones de venta distintas con un parámetro de comisión que supere los 999 de la tabla SALES, use el siguiente ejemplo. Este comando cuenta las comisiones superiores a 999 comprobando si el decimal está a más de 4 lugares del principio del valor de la comisión.

```
SELECT DISTINCT POSITION('.' IN commission), COUNT (POSITION('.' IN commission))
FROM sales
```

```
WHERE POSITION('.' IN commission) > 4
GROUP BY POSITION('.' IN commission)
ORDER BY 1,2;

+----------+-------+
|  position | count |
+----------+-------+
|         5 |   629 |
+----------+-------+
```

## Función QUOTE_IDENT

La función QUOTE_IDENT devuelve la cadena especificada como una cadena con comillas dobles al principio y al final. La salida de la función se puede utilizar como un identificador en una instrucción SQL. La función duplica de forma adecuada todas las comillas dobles integradas.

QUOTE_IDENT agrega comillas dobles solo donde sean necesarias para crear un identificador válido, si la cadena contiene caracteres que no son identificadores o que, de otro modo, se pasarían a minúsculas. Para devolver siempre una cadena con comillas simples, utilice QUOTE_LITERAL.

### Sintaxis

```
QUOTE_IDENT(string)
```

### Argumento

string

Una cadena CHAR o VARCHAR.

### Tipo de retorno

La función QUOTE_IDENT devuelve el mismo tipo de cadena que la cadena de entrada.

### Ejemplos

Para devolver la cadena "CAT" con comillas dobles, utilice el siguiente ejemplo.

```
SELECT QUOTE_IDENT('"CAT"');
```

```
+-------------+
| quote_ident |
+-------------+
| """CAT"""   |
+-------------+
```

En el siguiente ejemplo se utilizan datos de la tabla CATEGORY en la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para devolver la columna CATNAME entre comillas, use el siguiente ejemplo.

```
SELECT catid, QUOTE_IDENT(catname)
FROM category
ORDER BY 1,2;

+-------+-------------+
| catid | quote_ident |
+-------+-------------+
|     1 | "MLB"       |
|     2 | "NHL"       |
|     3 | "NFL"       |
|     4 | "NBA"       |
|     5 | "MLS"       |
|     6 | "Musicals"  |
|     7 | "Plays"     |
|     8 | "Opera"     |
|     9 | "Pop"       |
|    10 | "Jazz"      |
|    11 | "Classical" |
+-------+-------------+
```

## Función QUOTE_LITERAL

La función QUOTE_LITERAL devuelve la cadena específica como una cadena con comillas simples para que se puede utilizar como un literal de cadena en una instrucción SQL. Si el parámetro de entrada es un número, QUOTE_LITERAL lo trata como una cadena. Duplica de forma adecuada todas las comillas simples y las barras inversas integradas.

### Sintaxis

```
QUOTE_LITERAL(string)
```

Argumento

string

Una cadena CHAR o VARCHAR.

Tipo de retorno

La función QUOTE_LITERAL devuelve una cadena CHAR o VARCHAR que es el mismo tipo de datos que la cadena de entrada.

Ejemplos

Para devolver la cadena ''CAT'' con comillas SIMPLES, utilice el siguiente ejemplo.

```
SELECT QUOTE_LITERAL('''CAT''');

+---------------+
| quote_literal |
+---------------+
| '''CAT'''     |
+---------------+
```

En los siguientes ejemplos se utilizan datos de la tabla CATEGORY en la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para devolver la columna CATNAME entre comillas, use el siguiente ejemplo.

```
SELECT catid, QUOTE_LITERAL(catname)
FROM category
ORDER BY 1,2;

+-------+---------------+
| catid | quote_literal |
+-------+---------------+
|     1 | 'MLB'         |
|     2 | 'NHL'         |
|     3 | 'NFL'         |
|     4 | 'NBA'         |
|     5 | 'MLS'         |
|     6 | 'Musicals'    |
```

```
|     7 | 'Plays'       |
|     8 | 'Opera'       |
|     9 | 'Pop'         |
|    10 | 'Jazz'        |
|    11 | 'Classical'   |
+-------+---------------+
```

Para devolver la columna CATID entre comillas, use el siguiente ejemplo.

```
SELECT QUOTE_LITERAL(catid), catname
FROM category
ORDER BY 1,2;

+---------------+-----------+
| quote_literal |  catname  |
+---------------+-----------+
| '1'           | MLB       |
| '10'          | Jazz      |
| '11'          | Classical |
| '2'           | NHL       |
| '3'           | NFL       |
| '4'           | NBA       |
| '5'           | MLS       |
| '6'           | Musicals  |
| '7'           | Plays     |
| '8'           | Opera     |
| '9'           | Pop       |
+---------------+-----------+
```

## Función REGEXP_COUNT

Busca una cadena para un patrón de expresión regular y devuelve un valor entero que indica la cantidad de veces que el patrón especificado aparece en la cadena. Si no se encuentra coincidencia, la función devuelve `0`. Para obtener más información sobre las expresiones regulares, consulte Operadores POSIX y Expresión regular en la Wikipedia.

### Sintaxis

```
REGEXP_COUNT( source_string, pattern [, position [, parameters ] ] )
```

Argumentos

source_string

Una cadena CHAR o VARCHAR.

pattern

Un literal de cadena UTF-8 que representa un patrón de expresión regular. Para obtener más información, consulte Operadores POSIX.

position

(Opcional) Un número INTEGER que indica la posición dentro de source_string para comenzar la búsqueda. La posición se basa en la cantidad de caracteres, no bytes, por lo que los caracteres multibyte se cuentan como caracteres simples. El valor predeterminado es 1. Si el valor de posición es menor que 1, la búsqueda comienza en el primer carácter de source_string. Si el valor de posición es mayor que el número de caracteres de source_string, el resultado es 0.

parameters

(Opcional) Uno o varios literales de cadena que indican el grado de coincidencia de la función con el patrón. Los valores posibles son los siguientes:

- c: aplica la coincidencia que distingue entre mayúsculas y minúsculas. El comportamiento predeterminado es utilizar la coincidencia que distingue entre mayúsculas y minúsculas.
- i: aplica la coincidencia que no distingue entre mayúsculas y minúsculas.
- p: interpreta el patrón con el dialecto de expresión regular compatible con Perl (PCRE). Para obtener más información sobre PCRE, consulte Perl Compatible Regular Expressions en la Wikipedia.

Tipo de retorno

INTEGER

Ejemplos

Para contar la cantidad de veces en que aparece una secuencia de tres letras, use el siguiente ejemplo.

```
SELECT REGEXP_COUNT('abcdefghijklmnopqrstuvwxyz', '[a-z]{3}');

+--------------+
```

```
| regexp_count |
+--------------+
|            8 |
+--------------+
```

Para contar cuántas veces aparece la cadena FOX, con una coincidencia que no distingue entre mayúsculas y minúsculas, use el siguiente ejemplo.

```
SELECT REGEXP_COUNT('the fox', 'FOX', 1, 'i');

+--------------+
| regexp_count |
+--------------+
|            1 |
+--------------+
```

Para usar un patrón escrito en el dialecto de PCRE para localizar palabras que contengan al menos un número y una letra en minúsculas, use el siguiente ejemplo. El ejemplo utiliza el operador ?=, que tiene una connotación específica de anticipación en PCRE. En este ejemplo, se cuenta cuántas veces aparecen dichas palabras, con una coincidencia que distingue entre mayúsculas y minúsculas.

```
SELECT REGEXP_COUNT('passwd7 plain A1234 a1234', '(?=[^ ]*[a-z])(?=[^ ]*[0-9])[^ ]+',
 1, 'p');

+--------------+
| regexp_count |
+--------------+
|            2 |
+--------------+
```

Para usar un patrón escrito en el dialecto de PCRE para localizar palabras que contengan al menos un número y una letra en minúsculas, use el siguiente ejemplo. Se utiliza el operador ?=, que tiene una connotación específica en PCRE. En este ejemplo, se cuenta cuántas veces aparecen dichas palabras, pero difiere del ejemplo anterior, ya que se utiliza una coincidencia sin distinción entre mayúsculas y minúsculas.

```
SELECT REGEXP_COUNT('passwd7 plain A1234 a1234', '(?=[^ ]*[a-z])(?=[^ ]*[0-9])[^ ]+',
 1, 'ip');

+--------------+
```

```
| regexp_count |
+--------------+
|            3 |
+--------------+
```

En el siguiente ejemplo se utilizan datos de la tabla USERS en la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para contar la cantidad de veces en que el nombre del dominio de nivel superior es `org` o `edu`, use el siguiente ejemplo.

```
SELECT email, REGEXP_COUNT(email,'@[^.]*\.(org|edu)') FROM users
ORDER BY userid LIMIT 4;

+----------------------------------------------+--------------+
|                     email                    | regexp_count |
+----------------------------------------------+--------------+
| Etiam.laoreet.libero@sodalesMaurisblandit.edu |            1 |
| Suspendisse.tristique@nonnisiAenean.edu       |            1 |
| amet.faucibus.ut@condimentumegetvolutpat.ca   |            0 |
| sed@lacusUtnec.ca                             |            0 |
+----------------------------------------------+--------------+
```

## Función REGEXP_INSTR

Busca una cadena para un patrón de expresión regular y devuelve un valor entero que indica la posición de inicio o de finalización de la subcadena coincidente. Si no se encuentra coincidencia, la función devuelve `0`. REGEXP_INSTR es similar a la función POSITION, pero le permite buscar un patrón de expresión regular en una cadena. Para obtener más información sobre las expresiones regulares, consulte Operadores POSIX y Expresión regular en la Wikipedia.

### Sintaxis

```
REGEXP_INSTR( source_string, pattern [, position [, occurrence] [, option [, parameters
 ] ] ] ] )
```

### Argumentos

source_string

Una expresión de cadena, como un nombre de columna, que se buscará.

pattern

Un literal de cadena UTF-8 que representa un patrón de expresión regular. Para obtener más información, consulte Operadores POSIX.

position

(Opcional) Un número `INTEGER` que indica la posición dentro de source_string para comenzar la búsqueda. La posición se basa en la cantidad de caracteres, no bytes, por lo que los caracteres multibyte se cuentan como caracteres simples. El valor predeterminado es `1`. Si el valor de posición es menor que `1`, la búsqueda comienza en el primer carácter de source_string. Si el valor de posición es mayor que el número de caracteres de source_string, el resultado es `0`.

occurrence

(Opcional) Un número `INTEGER` positivo que indica qué coincidencia del patrón se va a utilizar. REGEXP_INSTR omite las primeras coincidencias *`occurrence`*`-1`. El valor predeterminado es `1`. Si occurrence es menor que `1` o mayor que el número de caracteres de source_string, la búsqueda se omite y el resultado es `0`.

option

(Opcional )Valor que indica si se va a devolver la posición del primer carácter de la coincidencia (`0`) o la posición del primer carácter situado a continuación del final de la coincidencia (`1`). Un valor distinto de cero es lo mismo que `1`. El valor predeterminado es `0`.

parameters

(Opcional) Uno o varios literales de cadena que indican el grado de coincidencia de la función con el patrón. Los valores posibles son los siguientes:

- c: aplica la coincidencia que distingue entre mayúsculas y minúsculas. El comportamiento predeterminado es utilizar la coincidencia que distingue entre mayúsculas y minúsculas.
- i: aplica la coincidencia que no distingue entre mayúsculas y minúsculas.
- e: extrae una subcadena mediante una subexpresión.

  Si pattern incluye una subexpresión, REGEXP_INSTR realiza la comparación con una subcadena utilizando la primera subexpresión de pattern. REGEXP_INSTR solo tiene en cuenta la primera subexpresión; las subexpresiones adicionales se omiten. Si el patrón no incluye una subexpresión, REGEXP_INSTR omite el parámetro 'e'.

- p: interpreta el patrón con el dialecto de expresión regular compatible con Perl (PCRE). Para obtener más información sobre PCRE, consulte Perl Compatible Regular Expressions en la Wikipedia.

Tipo de retorno

Entero

Ejemplos

En los siguientes ejemplos se utilizan datos de la tabla USERS en la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para buscar el carácter @ que comienza un nombre de dominio y devolver la posición inicial de la primera coincidencia, use el siguiente ejemplo.

```
SELECT email, REGEXP_INSTR(email, '@[^.]*')
FROM users
ORDER BY userid LIMIT 4;

+----------------------------------------------+--------------+
|                     email                    | regexp_instr |
+----------------------------------------------+--------------+
| Etiam.laoreet.libero@sodalesMaurisblandit.edu |           21 |
| Suspendisse.tristique@nonnisiAenean.edu       |           22 |
| amet.faucibus.ut@condimentumegetvolutpat.ca   |           17 |
| sed@lacusUtnec.ca                             |            4 |
+----------------------------------------------+--------------+
```

Para buscar variantes de la palabra `Center` y devolver la posición inicial de la primera coincidencia, use el siguiente ejemplo..

```
SELECT venuename, REGEXP_INSTR(venuename,'[cC]ent(er|re)$')
FROM venue
WHERE REGEXP_INSTR(venuename,'[cC]ent(er|re)$') > 0
ORDER BY venueid LIMIT 4;

+----------------------+--------------+
|       venuename      | regexp_instr |
+----------------------+--------------+
| The Home Depot Center |           16 |
| Izod Center           |            6 |
```

```
| Wachovia Center       |           10 |
| Air Canada Centre     |           12 |
+-----------------------+--------------+
```

Para encontrar la posición inicial de la primera vez que aparece la cadena FOX, con una lógica que no distingue entre mayúsculas y minúsculas, use el siguiente ejemplo.

```
SELECT REGEXP_INSTR('the fox', 'FOX', 1, 1, 0, 'i');

+--------------+
| regexp_instr |
+--------------+
|            5 |
+--------------+
```

Para usar un patrón escrito en el dialecto de PCRE para localizar palabras que contengan al menos un número y una letra en minúsculas, use el siguiente ejemplo. Se utiliza el operador ?=, que tiene una connotación específica de anticipación en PCRE. En este ejemplo, se encuentra la posición inicial de la segunda palabra que reúne esas características.

```
SELECT REGEXP_INSTR('passwd7 plain A1234 a1234', '(?=[^ ]*[a-z])(?=[^ ]*[0-9])[^ ]+',
 1, 2, 0, 'p');

+--------------+
| regexp_instr |
+--------------+
|           21 |
+--------------+
```

Para usar un patrón escrito en el dialecto de PCRE para localizar palabras que contengan al menos un número y una letra en minúsculas, use el siguiente ejemplo. Se utiliza el operador ?=, que tiene una connotación específica de anticipación en PCRE. En este ejemplo, se encuentra la posición inicial de la segunda palabra que reúne esas características, pero difiere del ejemplo anterior, ya que se utiliza una coincidencia sin distinción entre mayúsculas y minúsculas.

```
SELECT REGEXP_INSTR('passwd7 plain A1234 a1234', '(?=[^ ]*[a-z])(?=[^ ]*[0-9])[^ ]+',
 1, 2, 0, 'ip');

+--------------+
| regexp_instr |
```

```
+--------------+
|           15 |
+--------------+
```

## Función REGEXP_REPLACE

Busca una cadena para un patrón de expresión regular y reemplaza cada coincidencia del patrón con una cadena especificada. REGEXP_REPLACE es similar a Función REPLACE, pero le permite buscar un patrón de expresión regular en una cadena. Para obtener más información sobre las expresiones regulares, consulte Operadores POSIX y Expresión regular en la Wikipedia.

REGEXP_REPLACE es similar a Función TRANSLATE y a Función REPLACE, salvo que TRANSLATE realiza varias sustituciones de caracteres únicos y REPLACE sustituye una cadena entera por otra cadena, mientras que REGEXP_REPLACE le permite buscar un patrón de expresión regular en una cadena.

Sintaxis

```
REGEXP_REPLACE( source_string, pattern [, replace_string [ , position [, parameters
 ] ] ] )
```

Argumentos

source_string

Una expresión de cadena CHAR o VARCHAR, como un nombre de columna, que se buscará.

pattern

Un literal de cadena UTF-8 que representa un patrón de expresión regular. Para obtener más información, consulte Operadores POSIX.

replace_string

(Opcional) Una expresión de cadena CHAR o VARCHAR, como un nombre de columna, que reemplazará cada coincidencia del patrón. El valor predeterminado es una cadena vacía ( "" ).

position

(Opcional) Valor entero positivo que indica la posición dentro de source_string para comenzar la búsqueda. La posición se basa en la cantidad de caracteres, no bytes, por lo que los caracteres multibyte se cuentan como caracteres simples. El valor predeterminado es 1. Si el valor de posición es menor que 1, la búsqueda comienza en el primer carácter de source_string.

Si el valor de position (posición) es mayor que la cantidad de caracteres de source-string (cadena_de_origen), el resultado es source_string (cadena_de_origen).

parameters

(Opcional) Uno o varios literales de cadena que indican el grado de coincidencia de la función con el patrón. Los valores posibles son los siguientes:

- c: aplica la coincidencia que distingue entre mayúsculas y minúsculas. El comportamiento predeterminado es utilizar la coincidencia que distingue entre mayúsculas y minúsculas.
- i: aplica la coincidencia que no distingue entre mayúsculas y minúsculas.
- p: interpreta el patrón con el dialecto de expresión regular compatible con Perl (PCRE). Para obtener más información sobre PCRE, consulte Perl Compatible Regular Expressions en la Wikipedia.

Tipo de retorno

VARCHAR

Si patrón o replace_string es `NULL`, la función devuelve `NULL`.

Ejemplos

Para reemplazar todas las veces que aparece la cadena `FOX` en el valor `quick brown fox`, con una coincidencia que no distingue entre mayúsculas y minúsculas, use el siguiente ejemplo.

```
SELECT REGEXP_REPLACE('the fox', 'FOX', 'quick brown fox', 1, 'i');

+----------------------+
|   regexp_replace     |
+----------------------+
| the quick brown fox |
+----------------------+
```

En el siguiente ejemplo, se utiliza un patrón escrito en el dialecto de PCRE para localizar palabras que contengan al menos un número y una letra en minúsculas. Se utiliza el operador `?=`, que tiene una connotación específica de anticipación en PCRE. Para sustituir cada palabra que reúne esas características con el valor `[hidden]`, use el siguiente ejemplo.

```
SELECT REGEXP_REPLACE('passwd7 plain A1234 a1234', '(?=[^ ]*[a-z])(?=[^ ]*[0-9])[^ ]+',
 '[hidden]', 1, 'p');
```

```
+------------------------------+
|        regexp_replace        |
+------------------------------+
| [hidden] plain A1234 [hidden] |
+------------------------------+
```

En el siguiente ejemplo, se utiliza un patrón escrito en el dialecto de PCRE para localizar palabras que contengan al menos un número y una letra en minúsculas. Se utiliza el operador ?=, que tiene una connotación específica de anticipación en PCRE. Para reemplazar cada vez que aparece una palabra que reúne esas características con el valor [hidden], pero difiere del ejemplo anterior, ya que se utiliza una coincidencia sin distinción entre mayúsculas y minúsculas, use el siguiente ejemplo.

```
SELECT REGEXP_REPLACE('passwd7 plain A1234 a1234', '(?=[^ ]*[a-z])(?=[^ ]*[0-9])[^ ]+',
 '[hidden]', 1, 'ip');

+----------------------------------+
|          regexp_replace          |
+----------------------------------+
| [hidden] plain [hidden] [hidden] |
+----------------------------------+
```

En los siguientes ejemplos se utilizan datos de la tabla USERS en la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para eliminar @ y el nombre de dominio de direcciones de correo electrónico, use el siguiente ejemplo.

```
SELECT email, REGEXP_REPLACE(email, '@.*\\.(org|gov|com|edu|ca)$')
FROM users
ORDER BY userid LIMIT 4;

+-----------------------------------------------+----------------------+
|                     email                     |    regexp_replace    |
+-----------------------------------------------+----------------------+
| Etiam.laoreet.libero@sodalesMaurisblandit.edu | Etiam.laoreet.libero |
| Suspendisse.tristique@nonnisiAenean.edu       | Suspendisse.tristique |
| amet.faucibus.ut@condimentumegetvolutpat.ca   | amet.faucibus.ut     |
| sed@lacusUtnec.ca                             | sed                  |
+-----------------------------------------------+----------------------+
```

Para reemplazar los nombres de dominio de las direcciones de correo electrónico con `internal.company.com`, use el siguiente ejemplo.

```
SELECT email, REGEXP_REPLACE(email, '@.*\\.[[:alpha:]]{2,3}','@internal.company.com')
FROM users
ORDER BY userid LIMIT 4;

+----------------------------------------------
+-------------------------------------------+
|                     email                     |              regexp_replace
      |
+----------------------------------------------
+-------------------------------------------+
| Etiam.laoreet.libero@sodalesMaurisblandit.edu |
 Etiam.laoreet.libero@internal.company.com  |
| Suspendisse.tristique@nonnisiAenean.edu       |
 Suspendisse.tristique@internal.company.com |
| amet.faucibus.ut@condimentumegetvolutpat.ca   | amet.faucibus.ut@internal.company.com
      |
| sed@lacusUtnec.ca                             | sed@internal.company.com
      |
+----------------------------------------------
+-------------------------------------------+
```

## Función REGEXP_SUBSTR

Devuelve los caracteres de una cadena al buscar un patrón de expresión regular. REGEXP_SUBSTR es similar a la función Función SUBSTRING, pero le permite buscar un patrón de expresión regular en una cadena. Si la función no puede hacer coincidir la expresión regular con ningún carácter de la cadena, devuelve una cadena vacía. Para obtener más información sobre las expresiones regulares, consulte Operadores POSIX y Expresión regular en la Wikipedia.

Sintaxis

```
REGEXP_SUBSTR( source_string, pattern [, position [, occurrence [, parameters ] ] ] )
```

Argumentos

source_string

Una expresión de cadena que se va a buscar.

pattern

Un literal de cadena UTF-8 que representa un patrón de expresión regular. Para obtener más información, consulte Operadores POSIX.

position

Valor entero positivo que indica la posición dentro de source_string (cadena_de_origen) para comenzar la búsqueda. La posición se basa en la cantidad de caracteres, no bytes, por lo que los caracteres multibyte se cuentan como caracteres simples. El valor predeterminado de es 1. Si el valor de position (posición) es menor que 1, la búsqueda comienza en el primer carácter de source-string (cadena_de_origen). Si el valor de position (posición) es mayor que el número de caracteres de source-string (cadena_de_origen), el resultado es una cadena vacía ("").

occurrence

Un número entero positivo que indica qué coincidencia del patrón se va a utilizar. REGEXP_SUBSTR omite las primeras coincidencias especificadas por el valor de occurrence menos uno. El valor predeterminado de es 1. Si occurrence es menor que 1 o mayor que el número de caracteres de source_string, la búsqueda se omite y el resultado está vacío.

parameters

Uno o varios literales de cadena que indican el grado de coincidencia de la función con el patrón. Los valores posibles son los siguientes:

- c: aplica la coincidencia que distingue entre mayúsculas y minúsculas. El comportamiento predeterminado es utilizar la coincidencia que distingue entre mayúsculas y minúsculas.
- i: aplica la coincidencia que no distingue entre mayúsculas y minúsculas.
- e: extrae una subcadena mediante una subexpresión.

  Si pattern incluye una subexpresión, REGEXP_SUBSTR realiza la comparación con una subcadena utilizando la primera subexpresión de pattern. Una subexpresión es una expresión dentro del patrón que está entre paréntesis. Por ejemplo, para que el patrón `'This is a (\\w+)'` coincida con la primera expresión con la cadena `'This is a '` seguida de una palabra. En lugar de devolver el patrón, REGEXP_SUBSTR con el parámetro e devuelve solo la cadena dentro de la subexpresión.

  REGEXP_SUBSTR solo tiene en cuenta la primera subexpresión; las subexpresiones adicionales se omiten. Si el patrón no incluye una subexpresión, REGEXP_SUBSTR omite el parámetro 'e'.

- p: interpreta el patrón con el dialecto de expresión regular compatible con Perl (PCRE). Para obtener más información sobre PCRE, consulte Perl Compatible Regular Expressions en la Wikipedia.

Tipo de retorno

VARCHAR

Ejemplos

El siguiente ejemplo devuelve la parte de una dirección de correo electrónico entre el carácter @ y la extensión de dominio. Los datos de `users` consultados proceden de los datos de muestra de Amazon Redshift. Para obtener más información, consulte Base de datos de muestra.

```
SELECT email, regexp_substr(email,'@[^.]*')
FROM users
ORDER BY userid LIMIT 4;

                     email                      |      regexp_substr
------------------------------------------------+--------------------------
 Suspendisse.tristique@nonnisiAenean.edu        | @nonnisiAenean
 amet.faucibus.ut@condimentumegetvolutpat.ca    | @condimentumegetvolutpat
 sed@lacusUtnec.ca                              | @lacusUtnec
 Cum@accumsan.com                               | @accumsan
```

El siguiente ejemplo devuelve la parte de la entrada que corresponde a la primera vez que aparece la cadena F0X con una coincidencia que no distingue entre mayúsculas y minúsculas.

```
SELECT regexp_substr('the fox', 'FOX', 1, 1, 'i');

 regexp_substr
---------------
 fox
```

El siguiente ejemplo devuelve la parte de la entrada que corresponde a la segunda vez que aparece la cadena F0X con una coincidencia que no distingue entre mayúsculas y minúsculas. El resultado está vacío (no es nulo, longitud 0) porque no aparece por segunda vez.

```
SELECT regexp_substr('the fox', 'FOX', 1, 2, 'i');
```

```
 regexp_substr
---------------
```

El ejemplo siguiente devuelve la primera parte de la entrada que comienza en minúscula. Esto es funcionalmente idéntico a la misma instrucción SELECT sin el parámetro `c`.

```
SELECT regexp_substr('THE SECRET CODE IS THE LOWERCASE PART OF 1931abc0EZ.', '[a-z]+',
 1, 1, 'c');

 regexp_substr
---------------
 abc
```

En el siguiente ejemplo, se utiliza un patrón escrito en el dialecto de PCRE para localizar palabras que contengan al menos un número y una letra en minúsculas. Se utiliza el operador ?=, que tiene una connotación específica de anticipación en PCRE. En este ejemplo, se devuelve la parte de la entrada que corresponde a la segunda palabra que reúne esas características.

```
SELECT regexp_substr('passwd7 plain A1234 a1234', '(?=[^ ]*[a-z])(?=[^ ]*[0-9])[^ ]+',
 1, 2, 'p');

 regexp_substr
---------------
 a1234
```

En el siguiente ejemplo, se utiliza un patrón escrito en el dialecto de PCRE para localizar palabras que contengan al menos un número y una letra en minúsculas. Se utiliza el operador ?=, que tiene una connotación específica de anticipación en PCRE. En este ejemplo, se devuelve la parte de la entrada que corresponde a la segunda palabra que reúne esas características, pero difiere del ejemplo anterior, ya que se utiliza una coincidencia sin distinción entre mayúsculas y minúsculas.

```
SELECT regexp_substr('passwd7 plain A1234 a1234', '(?=[^ ]*[a-z])(?=[^ ]*[0-9])[^ ]+',
 1, 2, 'ip');

 regexp_substr
---------------
 A1234
```

En el ejemplo siguiente se utiliza una subexpresión para buscar la segunda cadena que coincida con el patrón `'this is a (\\w+)'` con una coincidencia que no distingue entre mayúsculas y minúsculas. Devuelve la subexpresión entre paréntesis.

```
SELECT regexp_substr(
               'This is a cat, this is a dog. This is a mouse.',
               'this is a (\\w+)', 1, 2, 'ie');

 regexp_substr
---------------
 dog
```

## Función REPEAT

Repite una cadena la cantidad especificada de veces. Si el parámetro de entrada es numérico, REPEAT lo trata como una cadena.

Sinónimo de Función REPLICATE.

Sintaxis

```
REPEAT(string, integer)
```

Argumentos

string

El primer parámetro de entrada es la cadena que se repetirá.

integer

El segundo parámetro es un valor `INTEGER` que indica la cantidad de veces que se repite la cadena.

Tipo de retorno

VARCHAR

Ejemplos

En el siguiente ejemplo se utilizan datos de la tabla CATEGORY en la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para repetir tres veces el valor de la columna CATID en la tabla CATEGORY, use el siguiente ejemplo.

```
SELECT catid, REPEAT(catid,3)
FROM category
ORDER BY 1,2;

+-------+--------+
| catid | repeat |
+-------+--------+
|     1 |    111 |
|     2 |    222 |
|     3 |    333 |
|     4 |    444 |
|     5 |    555 |
|     6 |    666 |
|     7 |    777 |
|     8 |    888 |
|     9 |    999 |
|    10 | 101010 |
|    11 | 111111 |
+-------+--------+
```

## Función REPLACE

Reemplaza todas las coincidencias de un conjunto de caracteres dentro de una cadena existente con otros caracteres especificados.

REPLACE es similar a Función TRANSLATE y a Función REGEXP_REPLACE, salvo que TRANSLATE realiza varias sustituciones de caracteres únicos y REGEXP_REPLACE le permite buscar un patrón de expresión regular en una cadena, mientras que REPLACE sustituye una cadena entera por otra cadena.

### Sintaxis

```
REPLACE(string, old_chars, new_chars)
```

### Argumentos

string

Cadena CHAR o VARCHAR que se buscará

old_chars

La cadena CHAR o VARCHAR que se va a sustituir.

new_chars

Nueva cadena CHAR o VARCHAR que sustituye old_string.

Tipo de retorno

VARCHAR

Si old_chars o new_chars es NULL, el valor devuelto es NULL.

Ejemplos

En el siguiente ejemplo se utilizan datos de la tabla CATEGORY en la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para convertir la cadena Shows en Theatre en el campo CATGROUP, use el siguiente ejemplo.

```
SELECT catid, catgroup, REPLACE(catgroup, 'Shows', 'Theatre')
FROM category
ORDER BY 1,2,3;

+-------+----------+----------+
| catid | catgroup | replace  |
+-------+----------+----------+
|     1 | Sports   | Sports   |
|     2 | Sports   | Sports   |
|     3 | Sports   | Sports   |
|     4 | Sports   | Sports   |
|     5 | Sports   | Sports   |
|     6 | Shows    | Theatre  |
|     7 | Shows    | Theatre  |
|     8 | Shows    | Theatre  |
|     9 | Concerts | Concerts |
|    10 | Concerts | Concerts |
|    11 | Concerts | Concerts |
+-------+----------+----------+
```

## Función REPLICATE

Sinónimo para la función REPEAT.

Consulte Función REPEAT.

## Función REVERSE

La función REVERSE opera en una cadena y devuelve los caracteres en orden inverso. Por ejemplo, `reverse('abcde')` devuelve `edcba`. Esta función trabaja sobre tipos de datos numéricos y de fecha, además de tipos de datos de caracteres; no obstante, en la mayoría de los casos, tiene valor práctico para las cadenas de caracteres.

Sintaxis

```
REVERSE( expression )
```

Argumento

expresión

Una expresión con un tipo de datos de carácter, fecha, marca temporal o número que representa el destino de la reversión de carácter. Todas las expresiones se convierten implícitamente en cadenas `VARCHAR`. Los espacios en blanco en las cadenas `CHAR` se ignoran.

Tipo de retorno

VARCHAR

Ejemplos

En los siguientes ejemplos se utilizan datos de las tablas USERS y SALES en la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para seleccionar cinco nombres distintos de ciudades y sus correspondientes nombres invertidos de la tabla USERS, use el siguiente ejemplo.

```
SELECT DISTINCT city AS cityname, REVERSE(cityname)
FROM users
ORDER BY city LIMIT 5;
```

```
+----------+----------+
| cityname | reverse  |
+----------+----------+
| Aberdeen | needrebA |
| Abilene  | enelibA  |
| Ada      | adA      |
| Agat     | tagA     |
| Agawam   | mawagA   |
+----------+----------+
```

Para seleccionar cinco ID de ventas y sus correspondientes ID invertidos vinculados convertidos a cadenas de caracteres, use el siguiente ejemplo.

```
SELECT salesid, REVERSE(salesid)
FROM sales
ORDER BY salesid DESC LIMIT 5;

+---------+---------+
| salesid | reverse |
+---------+---------+
|  172456 |  654271 |
|  172455 |  554271 |
|  172454 |  454271 |
|  172453 |  354271 |
|  172452 |  254271 |
+---------+---------+
```

## Función RTRIM

La función RTRIM recorta un conjunto especificado de caracteres desde el final de una cadena. Elimina la cadena más larga que contiene solo caracteres de la lista de caracteres de recorte. El recorte finaliza cuando no aparece ningún carácter de recorte en la cadena de entrada.

### Sintaxis

```
RTRIM( string, trim_chars )
```

### Argumentos

string

Una columna de cadena, una expresión o un literal de cadena que se va a recortar.

trim_chars

Es una columna de cadena, expresión o literal de cadena que representa los caracteres que se deben recortar desde el final de string. Si no se especifica, se utiliza un espacio como carácter de recorte.

Tipo de retorno

Cadena que es del mismo tipo de datos que el argumento string.

Ejemplo

En el siguiente ejemplo, se recortan espacios a la izquierda y a la derecha de la cadena `' abc '`:

```
select '     abc    ' as untrim, rtrim('     abc    ') as trim;

untrim    | trim
----------+------
   abc    |    abc
```

En el siguiente ejemplo, se eliminan las cadenas `'xyz'` a la derecha de la cadena `'xyzaxyzbxyzcxyz'`. Las coincidencias a la derecha de `'xyz'` se eliminan, pero las coincidencias internas dentro de la cadena no se eliminan.

```
select 'xyzaxyzbxyzcxyz' as untrim,
rtrim('xyzaxyzbxyzcxyz', 'xyz') as trim;

     untrim      |   trim
-----------------+-----------
 xyzaxyzbxyzcxyz | xyzaxyzbxyzc
```

En el siguiente ejemplo, se eliminan las partes a la derecha de la cadena `'setuphistorycassettes'` que coinciden con cualquiera de los caracteres de la lista trim_chars `'tes'`. Cualquier `t`, `e` o `s` que aparezca antes de cualquier carácter que no esté en la lista trim_chars al final de la cadena de entrada se eliminará.

```
SELECT rtrim('setuphistorycassettes', 'tes');

     rtrim
-----------------
```

```
 setuphistoryca
```

En el siguiente ejemplo, se recortan los caracteres "Park" del final de VENUENAME, cuando corresponde:

```
select venueid, venuename, rtrim(venuename, 'Park')
from venue
order by 1, 2, 3
limit 10;

venueid |         venuename          |          rtrim
--------+----------------------------+-------------------------
      1 | Toyota Park                | Toyota
      2 | Columbus Crew Stadium      | Columbus Crew Stadium
      3 | RFK Stadium                | RFK Stadium
      4 | CommunityAmerica Ballpark  | CommunityAmerica Ballp
      5 | Gillette Stadium           | Gillette Stadium
      6 | New York Giants Stadium    | New York Giants Stadium
      7 | BMO Field                  | BMO Field
      8 | The Home Depot Center      | The Home Depot Cente
      9 | Dick's Sporting Goods Park | Dick's Sporting Goods
     10 | Pizza Hut Park             | Pizza Hut
```

Tenga en cuenta que RTRIM elimina cualquiera de los caracteres `P`, `a`, `r` o `k` cuando aparecen al final de un VENUENAME.

## Función SOUNDEX

La función SOUNDEX devuelve el valor American Soundex que consiste en la primera letra de la cadena de entrada seguida de una codificación de 3 dígitos de los sonidos que representan la pronunciación en inglés de la cadena especificada. Por ejemplo, `Smith` y `Smyth` tienen el mismo valor de Soundex.

Sintaxis

```
SOUNDEX(string)
```

Argumentos

string

Especifica una cadena CHAR o VARCHAR que desea convertir en un valor de código American Soundex.

Tipo de retorno

VARCHAR(4)

Notas de uso

La función SOUNDEX solo convierte caracteres ASCII alfabéticos en minúsculas y mayúsculas en inglés, incluidas las letras a-z y A-Z. SOUNDEX omite otros caracteres. SOUNDEX devuelve un único valor Soundex para una cadena de varias palabras separadas por espacios.

```
SELECT SOUNDEX('AWS Amazon');

+---------+
| soundex |
+---------+
| A252    |
+---------+
```

SOUNDEX devuelve una cadena vacía si la cadena de entrada no contiene ninguna letra en inglés.

```
SELECT SOUNDEX('+-*/%');

+---------+
| soundex |
+---------+
|         |
+---------+
```

Ejemplos

Para devolver el valor de Soundex de Amazon, utilice el ejemplo siguiente.

```
SELECT SOUNDEX('Amazon');
```

```
+---------+
| soundex |
+---------+
| A525    |
+---------+
```

Para devolver el valor de Soundex de `smith` y `smyth`, utilice el ejemplo siguiente. Tenga en cuenta que los valores de Soundex son los mismos.

```
SELECT SOUNDEX('smith'), SOUNDEX('smyth');

+-------+-------+
| smith | smyth |
+-------+-------+
| S530  | S530  |
+-------+-------+
```

## Función SPLIT_PART

Divide una cadena en el delimitador especificado y devuelve la parte en la posición especificada.

Sintaxis

```
SPLIT_PART(string, delimiter, position)
```

Argumentos

string

Es una columna de cadena, una expresión o un literal de cadena que se va a dividir. La cadena puede ser CHAR o VARCHAR.

delimiter

Es la cadena delimitadora que indica las secciones del string de entrada.

Si el delimitador es un literal, enciérrelo entre comillas simples.

position

Posición de la porción de string a devolver (contando desde 1). Debe ser un número entero mayor que 0. Si position es mayor que la cantidad de porciones de la cadena, SPLIT_PART devuelve

una cadena vacía. Si no se encuentra el delimitador en cadena, entonces el valor devuelto contiene el contenido de la parte especificado, que podría ser la cadena completa o un valor vacío.

Tipo de retorno

Una cadena CHAR o VARCHAR, igual que el parámetro string.

Ejemplos

En el siguiente ejemplo, se divide un literal de cadena en partes mediante el uso del delimitador $ que devuelve la segunda parte.

```
select split_part('abc$def$ghi','$',2)

split_part
----------
def
```

En el siguiente ejemplo, se divide un literal de cadena en partes mediante el uso del delimitador $ que devuelve la segunda parte. Devuelve una cadena vacía porque no se encuentra la parte 4.

```
select split_part('abc$def$ghi','$',4)

split_part
----------

```

En el siguiente ejemplo, se divide un literal de cadena en partes mediante el uso del delimitador # que devuelve la segunda parte. Devuelve la cadena completa, que es la primera parte, porque no se encuentra el delimitador.

```
select split_part('abc$def$ghi','#',1)

split_part
------------
abc$def$ghi
```

En el siguiente ejemplo, se divide el campo de la marca temporal LISTTIME entre los componentes de año, mes y día.

```
select listtime, split_part(listtime,'-',1) as year,
split_part(listtime,'-',2) as month,
split_part(split_part(listtime,'-',3),' ',1) as day
from listing limit 5;

      listtime       | year | month | day
---------------------+------+-------+------
 2008-03-05 12:25:29 | 2008 | 03    | 05
 2008-09-09 08:03:36 | 2008 | 09    | 09
 2008-09-26 05:43:12 | 2008 | 09    | 26
 2008-10-04 02:00:30 | 2008 | 10    | 04
 2008-01-06 08:33:11 | 2008 | 01    | 06
```

En el siguiente ejemplo, se selecciona el campo de la marca temporal LISTTIME y se lo divide teniendo en cuenta el carácter `'-'` para obtener el mes (la segunda parte de la cadena LISTTIME). Luego, se cuenta la cantidad de entradas para cada mes:

```
select split_part(listtime,'-',2) as month, count(*)
from listing
group by split_part(listtime,'-',2)
order by 1, 2;

 month | count
-------+-------
    01 | 18543
    02 | 16620
    03 | 17594
    04 | 16822
    05 | 17618
    06 | 17158
    07 | 17626
    08 | 17881
    09 | 17378
    10 | 17756
    11 | 12912
    12 | 4589
```

## Función STRPOS

Devuelve la posición de una subcadena dentro de una cadena especificada.

Consulte Función CHARINDEX y Función POSITION para ver funciones similares.

### Sintaxis

```
STRPOS(string, substring )
```

### Argumentos

string

El primer parámetro de entrada es la cadena CHAR o VARCHAR que se buscará.

subcadena

El segundo parámetro es la subcadena que se va a buscar dentro de string (cadena).

### Tipo de retorno

INTEGER

La función STRPOS devuelve un valor INTEGER correspondiente a la posición de la subcadena (basado en uno, no basado en cero). La posición se basa en la cantidad de caracteres, no bytes, por lo que los caracteres multibyte se cuentan como caracteres simples.

### Notas de uso

STRPOS devuelve 0 si no se encuentra la subcadena dentro de cadena.

```
SELECT STRPOS('dogfish', 'fist');

+--------+
| strpos |
+--------+
|      0 |
+--------+
```

### Ejemplos

Para mostrar la posición de fish dentro de dogfish, utilice el ejemplo siguiente.

```
SELECT STRPOS('dogfish', 'fish');
```

```
+--------+
| strpos |
+--------+
|      4 |
+--------+
```

En el siguiente ejemplo se utilizan datos de la tabla SALES en la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para devolver la cantidad de transacciones de venta con un parámetro COMMISSION que supere los 999 de la tabla SALES, use el siguiente ejemplo.

```
SELECT DISTINCT STRPOS(commission, '.'),
COUNT (STRPOS(commission, '.'))
FROM sales
WHERE STRPOS(commission, '.') > 4
GROUP BY STRPOS(commission, '.')
ORDER BY 1, 2;

+--------+-------+
| strpos | count |
+--------+-------+
|      5 |   629 |
+--------+-------+
```

## Función STRTOL

Convierte una expresión de cadena de un número de la base especificada al valor entero equivalente. El valor convertido debe estar dentro del rango señalado de 64 bits.

### Sintaxis

```
STRTOL(num_string, base)
```

### Argumentos

num_string

La expresión de cadena de un número que se convertirá. Si num_string está vacío (`''`) o comienza con el carácter nulo (`'\0'`), el valor convertido es `0`. Si num_string es una columna que tiene un valor NULL, STRTOL devuelve `NULL`. La cadena puede comenzar con cualquier

cantidad de espacio en blanco, opcionalmente seguido de un signo más "+" o menos "-" único para indicar positivo o negativo. El valor predeterminado es '+'. Si base es 16, la cadena puede comenzar opcionalmente por "0x".

base

INTEGER entre 2 y 36.

Tipo de retorno

BIGINT

Si num_string es nulo, la función devuelve NULL.

Ejemplos

Para convertir pares de cadenas y valores base a valores enteros, use los ejemplos siguientes.

```
SELECT STRTOL('0xf',16);

+--------+
| strtol |
+--------+
|     15 |
+--------+

SELECT STRTOL('abcd1234',16);

+------------+
|   strtol   |
+------------+
| 2882343476 |
+------------+

SELECT STRTOL('1234567', 10);

+---------+
| strtol  |
+---------+
| 1234567 |
+---------+
```

```
SELECT STRTOL('1234567', 8);

+--------+
| strtol |
+--------+
| 342391 |
+--------+

SELECT STRTOL('110101', 2);

+--------+
| strtol |
+--------+
|     53 |
+--------+

SELECT STRTOL('\0', 2);

+--------+
| strtol |
+--------+
|      0 |
+--------+
```

## Función SUBSTRING

Devuelve el subconjunto de una cadena basado en la posición inicial especificada.

Si la entrada es una cadena de caracteres, la posición inicial y el número de caracteres extraídos se basan en caracteres, y no en bytes, de modo tal que los caracteres de varios bytes se cuentan como si fueran simples. Si la entrada es una expresión binaria, la posición inicial y la subcadena extraída se basan en bytes. No puede especificar una longitud negativa, pero puede especificar una posición de inicio negativa.

Sintaxis

```
SUBSTRING(character_string FROM start_position [ FOR number_characters ] )
```

```
SUBSTRING(character_string, start_position, number_characters )
```

```
SUBSTRING(binary_expression, start_byte, number_bytes )
```

```
SUBSTRING(binary_expression, start_byte )
```

Argumentos

character_string

La cadena que se buscará. Los tipos de datos que no son caracteres se tratan como una cadena.

start_position

La posición dentro de la cadena para comenzar la extracción, comenzando en 1. En valor de start_position (posición_de_inicio) se basa en la cantidad de caracteres, no bytes, por lo que los caracteres multibyte se cuentan como caracteres simples. Este número puede ser negativo.

number_characters

La cantidad de caracteres para extraer (la longitud de la subcadena). El valor de number_characters (número_de_caracteres) se basa en la cantidad de caracteres, no bytes, por lo que los caracteres multibyte se cuentan como caracteres simples. Este número no puede ser negativo.

binary_expression

La expresión_binaria del tipo de datos VARBYTE que se buscará.

start_byte

La posición dentro de la expresión binaria desde donde comienza la extracción, con punto de partida en 1. Este número puede ser negativo.

number_bytes

La cantidad de bytes para extraer, es decir, la longitud de la subcadena. Este número no puede ser negativo.

Tipo de retorno

VARCHAR o VARBYTE en función de la entrada.

Notas de uso

A continuación, encontrará algunos ejemplos de cómo puede utilizar start_position y number_characters para extraer subcadenas desde varias posiciones de una cadena.

El siguiente ejemplo devuelve una cadena de cuatro caracteres comenzando con el sexto carácter.

```
select substring('caterpillar',6,4);
substring
-----------
pill
(1 row)
```

Si start_position + number_characters supera la longitud de la cadena, SUBSTRING devuelve una subcadena que comienza en start_position y llega hasta el final de la cadena. Por ejemplo:

```
select substring('caterpillar',6,8);
substring
-----------
pillar
(1 row)
```

Si `start_position` es negativo o 0, la función SUBSTRING devuelve una cadena que comienza en el primer carácter de la cadena con una longitud de `start_position` + `number_characters` -1. Por ejemplo:

```
select substring('caterpillar',-2,6);
substring
-----------
cat
(1 row)
```

Si `start_position` + `number_characters` -1 es menor o igual a cero, SUBSTRING devuelve una cadena vacía. Por ejemplo:

```
select substring('caterpillar',-5,4);
substring
-----------

(1 row)
```

Ejemplos

El siguiente ejemplo devuelve el mes de la cadena LISTTIME en la tabla LISTING:

```
select listid, listtime,
```

```
substring(listtime, 6, 2) as month
from listing
order by 1, 2, 3
limit 10;

 listid |      listtime       | month
--------+---------------------+-------
      1 | 2008-01-24 06:43:29 | 01
      2 | 2008-03-05 12:25:29 | 03
      3 | 2008-11-01 07:35:33 | 11
      4 | 2008-05-24 01:18:37 | 05
      5 | 2008-05-17 02:29:11 | 05
      6 | 2008-08-15 02:08:13 | 08
      7 | 2008-11-15 09:38:15 | 11
      8 | 2008-11-09 05:07:30 | 11
      9 | 2008-09-09 08:03:36 | 09
     10 | 2008-06-17 09:44:54 | 06
(10 rows)
```

El siguiente ejemplo es igual al anterior, pero utiliza la opción FROM...FOR:

```
select listid, listtime,
substring(listtime from 6 for 2) as month
from listing
order by 1, 2, 3
limit 10;

 listid |      listtime       | month
--------+---------------------+-------
      1 | 2008-01-24 06:43:29 | 01
      2 | 2008-03-05 12:25:29 | 03
      3 | 2008-11-01 07:35:33 | 11
      4 | 2008-05-24 01:18:37 | 05
      5 | 2008-05-17 02:29:11 | 05
      6 | 2008-08-15 02:08:13 | 08
      7 | 2008-11-15 09:38:15 | 11
      8 | 2008-11-09 05:07:30 | 11
      9 | 2008-09-09 08:03:36 | 09
     10 | 2008-06-17 09:44:54 | 06
(10 rows)
```

No se puede utilizar SUBSTRING para extraer de forma predecible el prefijo de una cadena que pueda contener caracteres de varios bytes, ya que es necesario especificar la longitud de una

cadena de varios bytes en función de la cantidad de bytes, y no de la cantidad de caracteres. Para extraer el segmento inicial de una cadena en función de la longitud en bytes, puede utilizar CAST y convertir la cadena en VARCHAR(byte_length) para truncarla, donde byte_length es la longitud requerida. En el siguiente ejemplo, se extraen los 5 primeros bytes de la cadena `'Fourscore and seven'`.

```
select cast('Fourscore and seven' as varchar(5));

varchar
-------
Fours
```

En el siguiente ejemplo, se muestra una posición inicial negativa de un valor binario `abc`. Dado que la posición inicial es -3, la subcadena se extrae desde el principio del valor binario. El resultado se muestra automáticamente como representación hexadecimal de la subcadena binaria.

```
select substring('abc'::varbyte, -3);

 substring
-----------
 616263
```

En el siguiente ejemplo, se muestra un 1 en la posición inicial de un valor binario `abc`. Debido a que no se especifica la longitud, la cadena se extrae desde la posición inicial hasta el final de la cadena. El resultado se muestra automáticamente como representación hexadecimal de la subcadena binaria.

```
select substring('abc'::varbyte, 1);

 substring
-----------
 616263
```

En el siguiente ejemplo, se muestra un 3 en la posición inicial de un valor binario `abc`. Debido a que no se especifica la longitud, la cadena se extrae desde la posición inicial hasta el final de la cadena. El resultado se muestra automáticamente como representación hexadecimal de la subcadena binaria.

```
select substring('abc'::varbyte, 3);
```

```
 substring
-----------
 63
```

En el siguiente ejemplo, se muestra un 2 en la posición inicial de un valor binario `abc`. La cadena se extrae desde la posición inicial hasta la posición 10; no obstante, el final de la cadena se encuentra en la posición 3. El resultado se muestra automáticamente como representación hexadecimal de la subcadena binaria.

```
select substring('abc'::varbyte, 2, 10);

 substring
-----------
 6263
```

En el siguiente ejemplo, se muestra un 2 en la posición inicial de un valor binario `abc`. La cadena se extrae a partir de la posición inicial de 1 byte. El resultado se muestra automáticamente como representación hexadecimal de la subcadena binaria.

```
select substring('abc'::varbyte, 2, 1);

 substring
-----------
 62
```

El ejemplo siguiente devuelve el nombre `Ana` que aparece después del último espacio de la cadena de entrada `Silva, Ana`.

```
select reverse(substring(reverse('Silva, Ana'), 1, position(' ' IN reverse('Silva,
 Ana'))))

 reverse
-----------
 Ana
```

## Función TEXTLEN

Sinónimo de la función LEN.

Consulte Función LEN.

## Función TRANSLATE

Para una expresión dada, reemplaza todas las coincidencias de caracteres especificados con sustitutos especificados. Los caracteres existentes se asignan a caracteres de reemplazo en función de su posición en los argumentos characters_to_replace y characters_to_substitute. Si se especifican más caracteres en el argumento characters_to_replace que en el argumento characters_to_substitute, los caracteres adicionales del argumento characters_to_replace se omiten en el valor devuelto.

TRANSLATE es similar a Función REPLACE y a Función REGEXP_REPLACE, salvo que REPLACE sustituye una cadena entera por otra cadena y REGEXP_REPLACE le permite buscar un patrón de expresión regular en una cadena para, mientras que TRANSLATE realiza varias sustituciones de caracteres únicos.

Si un argumento es nulo, el valor de retorno es NULL.

### Sintaxis

```
TRANSLATE( expression, characters_to_replace, characters_to_substitute )
```

### Argumentos

expresión

La expresión que se traducirá.

characters_to_replace

Una cadena que tiene los caracteres que se reemplazarán.

characters_to_substitute

Una cadena que tiene los caracteres que se sustituirán.

### Tipo de retorno

VARCHAR

### Ejemplos

Para sustituir varios caracteres en una cadena, use el siguiente ejemplo.

```
SELECT TRANSLATE('mint tea', 'inea', 'osin');
```

```
+-----------+
| translate |
+-----------+
| most tin  |
+-----------+
```

En los siguientes ejemplos se utilizan datos de la tabla USERS en la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para reemplazar el signo (@) con un punto para todos los valores en una columna, use el siguiente ejemplo.

```
SELECT email, TRANSLATE(email, '@', '.') as obfuscated_email
FROM users LIMIT 10;

+--------------------------------------+--------------------------------------+
|                email                 |          obfuscated_email            |
+--------------------------------------+--------------------------------------+
| Cum@accumsan.com                     | Cum.accumsan.com                     |
| lorem.ipsum@Vestibulumante.com       | lorem.ipsum.Vestibulumante.com       |
| non.justo.Proin@ametconsectetuer.edu | non.justo.Proin.ametconsectetuer.edu |
| non.ante.bibendum@porttitortellus.org | non.ante.bibendum.porttitortellus.org |
| eros@blanditatnisi.org               | eros.blanditatnisi.org               |
| augue@Donec.ca                       | augue.Donec.ca                       |
| cursus@pedeacurna.edu                | cursus.pedeacurna.edu                |
| at@Duis.com                          | at.Duis.com                          |
| quam@facilisisvitaeorci.ca           | quam.facilisisvitaeorci.ca           |
| mi.lorem@nunc.edu                    | mi.lorem.nunc.edu                    |
+--------------------------------------+--------------------------------------+
```

Para reemplazar espacios con guiones bajos y quitar los puntos para todos los valores en una columna, use el siguiente ejemplo.

```
SELECT city, TRANSLATE(city, ' .', '_')
FROM users
WHERE city LIKE 'Sain%' OR city LIKE 'St%'
GROUP BY city
ORDER BY city;

+----------------+---------------+
|      city      |   translate   |
```

```
+----------------+---------------+
| Saint Albans   | Saint_Albans  |
| Saint Cloud    | Saint_Cloud   |
| Saint Joseph   | Saint_Joseph  |
| Saint Louis    | Saint_Louis   |
| Saint Paul     | Saint_Paul    |
| St. George     | St_George     |
| St. Marys      | St_Marys      |
| St. Petersburg | St_Petersburg |
| Stafford       | Stafford      |
| Stamford       | Stamford      |
| Stanton        | Stanton       |
| Starkville     | Starkville    |
| Statesboro     | Statesboro    |
| Staunton       | Staunton      |
| Steubenville   | Steubenville  |
| Stevens Point  | Stevens_Point |
| Stillwater     | Stillwater    |
| Stockton       | Stockton      |
| Sturgis        | Sturgis       |
+----------------+---------------+
```

## Función TRIM

Recorta una cadena mediante los espacios en blanco o los caracteres específicos.

### Sintaxis

```
TRIM( [ BOTH | LEADING | TRAILING ] [trim_chars FROM ] string )
```

### Argumentos

BOTH | LEADING | TRAILING

(Opcional) Especifica el lugar desde el que se van a recortar los caracteres. Utilice BOTH para eliminar los caracteres iniciales y finales, utilice LEADING para eliminar solo los caracteres principales y utilice TRAILING para eliminar solo los caracteres finales. Si se omite este parámetro, se recortan los caracteres iniciales y finales.

trim_chars

(Opcional) Los caracteres que se recortarán de la cadena. Si se omite este parámetro, se recortan los espacios en blanco.

string

La cadena que se recortará.

Tipo de retorno

La función TRIM devuelve una cadena VARCHAR o CHAR. Si utiliza la función TRIM con un comando SQL, Amazon Redshift convierte de forma implícita los resultados a VARCHAR. Si utiliza la función TRIM en la lista SELECT para una función SQL, Amazon Redshift no convierte de forma implícita los resultados y es posible que necesite realizar una conversión explícita para evitar un error de incompatibilidad con el tipo de dato. Consulte las funciones Función CAST y Función CONVERT para obtener información acerca de conversiones explícitas.

Ejemplos

Para recortar espacios a la izquierda y a la derecha de la cadena  dog , use el siguiente ejemplo.

```
SELECT TRIM('    dog    ');

+-------+
| btrim |
+-------+
| dog   |
+-------+
```

Para recortar espacios a la izquierda y a la derecha de la cadena  dog , use el siguiente ejemplo.

```
SELECT TRIM(BOTH FROM '    dog    ');

+-------+
| btrim |
+-------+
| dog   |
+-------+
```

Para eliminar las comillas dobles de inicio de la cadena "dog", utilice el siguiente ejemplo.

```
SELECT TRIM(LEADING '"' FROM'"dog"');

+-------+
| ltrim |
```

```
+-------+
| dog"  |
+-------+
```

Para eliminar las dobles comillas finales de la cadena ''dog'', utilice el siguiente ejemplo.

```
SELECT TRIM(TRAILING '"' FROM'"dog"');

+-------+
| rtrim |
+-------+
| "dog  |
+-------+
```

TRIM elimina cualquiera de los caracteres de trim_chars cuando aparecen al principio o al final de cadena. En el siguiente ejemplo, se recortan los caracteres “C”, “D” y “G” cuando aparecen al principio o al final de VENUENAME, que es una columna VARCHAR. Para obtener más información, consulte Base de datos de muestra.

```
SELECT venueid, venuename, TRIM('CDG' FROM venuename)
FROM venue
WHERE venuename LIKE '%Park'
ORDER BY 2
LIMIT 7;

+---------+----------------------------+----------------------------+
| venueid |         venuename          |           btrim            |
+---------+----------------------------+----------------------------+
|     121 | AT&T Park                  | AT&T Park                  |
|     109 | Citizens Bank Park         | itizens Bank Park          |
|     102 | Comerica Park              | omerica Park               |
|       9 | Dick's Sporting Goods Park | ick's Sporting Goods Park  |
|      97 | Fenway Park                | Fenway Park                |
|     112 | Great American Ball Park   | reat American Ball Park    |
|     114 | Miller Park                | Miller Park                |
+---------+----------------------------+----------------------------+
```

## Función UPPER

Convierte una cadena a mayúsculas. UPPER admite caracteres multibyte UTF-8 de hasta un máximo de cuatro bytes por carácter.

### Sintaxis

```
UPPER(string)
```

### Argumentos

string

El parámetro de entrada es una cadena VARCHAR o cualquier otro tipo de datos, como CHAR, que se pueda convertir de forma implícita a VARCHAR.

### Tipo de retorno

La función UPPER devuelve una cadena de caracteres que tiene el mismo tipo de datos que la cadena de entrada. Por ejemplo, la función devolverá una cadena VARCHAR si la entrada es una cadena VARCHAR.

### Ejemplos

En el siguiente ejemplo se utilizan datos de la tabla CATEGORY en la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

Para convertir el campo CATNAME a mayúsculas, use lo siguiente.

```
SELECT catname, UPPER(catname)
FROM category
ORDER BY 1,2;

+-----------+-----------+
|  catname  |   upper   |
+-----------+-----------+
| Classical | CLASSICAL |
| Jazz      | JAZZ      |
| MLB       | MLB       |
| MLS       | MLS       |
| Musicals  | MUSICALS  |
| NBA       | NBA       |
| NFL       | NFL       |
| NHL       | NHL       |
| Opera     | OPERA     |
| Plays     | PLAYS     |
```

```
| Pop        | POP        |
+------------+------------+
```

# Funciones de información acerca del tipo SUPER

A continuación, puede encontrar una descripción de las funciones de información acerca de los tipos para SQL que son compatibles con Amazon Redshift para obtener información dinámica de las entradas con el tipo de datos SUPER.

Temas


- Función DECIMAL_PRECISION
- Función DECIMAL_SCALE
- Función IS_ARRAY
- Función IS_BIGINT
- Función IS_BOOLEAN
- Función IS_CHAR
- Función IS_DECIMAL
- Función IS_FLOAT
- Función IS_INTEGER
- Función IS_OBJECT
- Función IS_SCALAR
- Función IS_SMALLINT
- Función IS_VARCHAR
- Función JSON_SIZE
- Función JSON_TYPEOF
- SIZE


## Función DECIMAL_PRECISION

Verifica la precisión del número total máximo de dígitos decimales que se almacenarán. Este número incluye los dígitos a la izquierda y a la derecha de la coma decimal. El rango de precisión es de 1 a 38, con un valor predeterminado de 38.

### Sintaxis

```
DECIMAL_PRECISION(super_expression)
```

### Argumentos

super_expression

Una expresión o columna SUPER.

### Tipo de retorno

INTEGER

### Ejemplos

Para aplicar la función DECIMAL_PRECISION a la tabla t, use el siguiente ejemplo.

```
CREATE TABLE t(s SUPER);

INSERT INTO t VALUES (3.14159);

SELECT DECIMAL_PRECISION(s) FROM t;

+-------------------+
| decimal_precision |
+-------------------+
|                 6 |
+-------------------+
```

## Función DECIMAL_SCALE

Verifica el número de dígitos decimales que se almacenarán a la derecha de la coma decimal. El rango de la escala va de 0 al punto de precisión, con un valor predeterminado de 0.

### Sintaxis

```
DECIMAL_SCALE(super_expression)
```

### Argumentos

super_expression

Una expresión o columna SUPER.

### Tipo de retorno

INTEGER

### Ejemplos

Para aplicar la función DECIMAL_SCALE a la tabla t, use el siguiente ejemplo.

```
CREATE TABLE t(s SUPER);

INSERT INTO t VALUES (3.14159);

SELECT DECIMAL_SCALE(s) FROM t;

+---------------+
| decimal_scale |
+---------------+
|             5 |
+---------------+
```

## Función IS_ARRAY

Verifica si una variable está en una matriz. La función devuelve `true` si la variable es una matriz. La función también incluye matrices vacías. De lo contrario, la función devuelve `false` para todos los demás valores, incluido el valor nulo.

### Sintaxis

```
IS_ARRAY(super_expression)
```

### Argumentos

super_expression

Una expresión o columna SUPER.

Tipo de retorno

BOOLEAN

Ejemplos

Para comprobar si [1,2] es una matriz que utiliza la función IS_ARRAY, utilice el siguiente ejemplo.

```
SELECT IS_ARRAY(JSON_PARSE('[1,2]'));

+----------+
| is_array |
+----------+
| true     |
+----------+
```

## Función IS_BIGINT

Verifica si un valor es BIGINT. La función IS_BIGINT devuelve `true` para los números de la escala 0 en el rango de 64 bits. De lo contrario, la función devuelve `false` para todos los demás valores, incluidos el valor nulo y los números con coma flotante.

La función IS_BIGINT es un superconjunto de IS_INTEGER.

Sintaxis

```
IS_BIGINT(super_expression)
```

Argumentos

super_expression

Una expresión o columna SUPER.

Tipo de retorno

BOOLEAN

Ejemplos

Para comprobar si 5 es BIGINT con la función IS_BIGINT, utilice el siguiente ejemplo.

```
CREATE TABLE t(s SUPER);

INSERT INTO t VALUES (5);

SELECT s, IS_BIGINT(s) FROM t;

+---+-----------+
| s | is_bigint |
+---+-----------+
| 5 | true      |
+---+-----------+
```

## Función IS_BOOLEAN

Verifica si un valor es BOOLEAN. La función IS_BOOLEAN devuelve `true` para los booleanos JSON constantes. La función devuelve `false` para todos los demás valores, incluido el valor nulo.

### Sintaxis

```
IS_BOOLEAN(super_expression)
```

### Argumentos

super_expression

Una expresión o columna SUPER.

### Tipo de retorno

BOOLEAN

### Ejemplos

Para comprobar si TRUE es BOOLEAN con la función IS_BOOLEAN, utilice el siguiente ejemplo.

```
CREATE TABLE t(s SUPER);

INSERT INTO t VALUES (TRUE);

SELECT s, IS_BOOLEAN(s) FROM t;
```

```
+------+------------+
|  s   | is_boolean |
+------+------------+
| true | true       |
+------+------------+
```

## Función IS_CHAR

Verifica si un valor es CHAR. La función IS_CHAR devuelve `true` para las cadenas que solo tienen caracteres ASCII, ya que el tipo CHAR solo puede almacenar caracteres con formato ASCII. La función devuelve `false` para todos los demás valores.

### Sintaxis

```
IS_CHAR(super_expression)
```

### Argumentos

super_expression

Una expresión o columna SUPER.

### Tipo de retorno

BOOLEAN

### Ejemplos

Para comprobar si `t` es CHAR con la función IS_CHAR, utilice el siguiente ejemplo.

```
CREATE TABLE t(s SUPER);

INSERT INTO t VALUES ('t');

SELECT s, IS_CHAR(s) FROM t;

+-----+---------+
|  s  | is_char |
+-----+---------+
```

```
| "t" | true    |
+-----+---------+
```

## Función IS_DECIMAL

Verifica si un valor es `DECIMAL`. La función IS_DECIMAL devuelve `true` para los números que no sean de coma flotante. La función devuelve `false` para todos los demás valores, incluido el valor nulo.

La función IS_DECIMAL es un superconjunto de IS_BIGINT.

### Sintaxis

```
IS_DECIMAL(super_expression)
```

### Argumentos

super_expression

Una expresión o columna `SUPER`.

### Tipo de retorno

`BOOLEAN`

### Ejemplos

Para comprobar si `1.22` es `DECIMAL` con la función IS_DECIMAL, utilice el siguiente ejemplo.

```
CREATE TABLE t(s SUPER);

INSERT INTO t VALUES (1.22);

SELECT s, IS_DECIMAL(s) FROM t;

+------+------------+
|  s   | is_decimal |
+------+------------+
| 1.22 | true       |
+------+------------+
```

## Función IS_FLOAT

Verifica si un valor es un número de coma flotante. La función IS_FLOAT devuelve `true` para los números de coma flotante (`FLOAT4` y `FLOAT8`). La función devuelve `false` para todos los demás valores.

El conjunto de IS_DECIMAL y el conjunto de IS_FLOAT son discontinuos.

### Sintaxis

```
IS_FLOAT(super_expression)
```

### Argumentos

super_expression

Una expresión o columna `SUPER`.

### Tipo de retorno

`BOOLEAN`

### Ejemplos

Para comprobar si `2.22::FLOAT` es `FLOAT` con la función IS_FLOAT, utilice el siguiente ejemplo.

```
CREATE TABLE t(s SUPER);

INSERT INTO t VALUES(2.22::FLOAT);

SELECT s, IS_FLOAT(s) FROM t;

+---------+----------+
|    s    | is_float |
+---------+----------+
| 2.22e+0 | true     |
+---------+----------+
```

## Función IS_INTEGER

Devuelve `true` para los números de la escala 0 en el rango de 32 bits y `false` para todo lo demás (incluidos los valores nulos y los números de coma flotante).

La función IS_INTEGER es un superconjunto de la función IS_SMALLINT.

Sintaxis

```
IS_INTEGER(super_expression)
```

Argumentos

super_expression

Una expresión o columna SUPER.

Tipo de retorno

BOOLEAN

Ejemplos

Para comprobar si 5 es INTEGER con la función IS_INTEGER, utilice el siguiente ejemplo.

```
CREATE TABLE t(s SUPER);

INSERT INTO t VALUES (5);

SELECT s, IS_INTEGER(s) FROM t;

+---+------------+
| s | is_integer |
+---+------------+
| 5 | true       |
+---+------------+
```

## Función IS_OBJECT

Verifica si una variable es un objeto. La función IS_OBJECT devuelve `true` para los objetos, incluidos los objetos vacíos. La función devuelve `false` para todos los demás valores, incluido el valor nulo.

Sintaxis

```
IS_OBJECT(super_expression)
```

### Argumentos

super_expression

Una expresión o columna SUPER.

### Tipo de retorno

BOOLEAN

### Ejemplos

Para comprobar si `{"name": "Joe"}` es un objeto con la función IS_OBJECT, utilice el siguiente ejemplo.

```
CREATE TABLE t(s super);

INSERT INTO t VALUES (JSON_PARSE('{"name": "Joe"}'));

SELECT s, IS_OBJECT(s) FROM t;

+----------------+-----------+
|       s        | is_object |
+----------------+-----------+
| {"name":"Joe"} | true      |
+----------------+-----------+
```

## Función IS_SCALAR

Verifica si una variable es escalar. La función IS_SCALAR devuelve `true` para cualquier valor que no sea una matriz o un objeto. La función devuelve `false` para todos los demás valores, incluido el valor nulo.

El conjunto de IS_ARRAY, IS_OBJECT e IS_SCALAR cubre todos los valores, excepto los nulos.

### Sintaxis

```
IS_SCALAR(super_expression)
```

### Argumentos

super_expression

Una expresión o columna SUPER.

### Tipo de retorno

BOOLEAN

### Ejemplos

Para comprobar si `{"name": "Joe"}` es escalar con la función IS_SCALAR, utilice el siguiente ejemplo.

```
CREATE TABLE t(s SUPER);

INSERT INTO t VALUES (JSON_PARSE('{"name": "Joe"}'));

SELECT s, IS_SCALAR(s.name) FROM t;

+----------------+-----------+
|       s        | is_scalar |
+----------------+-----------+
| {"name":"Joe"} | true      |
+----------------+-----------+
```

## Función IS_SMALLINT

Verifica si una variable es SMALLINT. La función IS_SMALLINT devuelve `true` para los números de la escala 0 en el rango de 16 bits. La función devuelve `false` para todos los demás valores, incluidos los valores nulos y los números con coma flotante.

### Sintaxis

```
IS_SMALLINT(super_expression)
```

### Argumentos

super_expression

Una expresión o columna SUPER.

Return

BOOLEAN

Ejemplos

Para comprobar si 5 es SMALLINT con la función IS_SMALLINT, utilice el siguiente ejemplo.

```
CREATE TABLE t(s SUPER);

INSERT INTO t VALUES (5);

SELECT s, IS_SMALLINT(s) FROM t;

+---+-------------+
| s | is_smallint |
+---+-------------+
| 5 | true        |
+---+-------------+
```

## Función IS_VARCHAR

Verifica si una variable es VARCHAR. La función IS_VARCHAR devuelve `true` para todas las cadenas. La función devuelve `false` para todos los demás valores.

La función IS_VARCHAR es un superconjunto de la función IS_CHAR.

Sintaxis

```
IS_VARCHAR(super_expression)
```

Argumentos

super_expression

Una expresión o columna SUPER.

Tipo de retorno

BOOLEAN

### Ejemplos

Para comprobar si `abc` es `VARCHAR` con la función IS_VARCHAR, utilice el siguiente ejemplo.

```
CREATE TABLE t(s SUPER);

INSERT INTO t VALUES ('abc');

SELECT s, IS_VARCHAR(s) FROM t;

+-------+------------+
|   s   | is_varchar |
+-------+------------+
| "abc" | true       |
+-------+------------+
```

## Función JSON_SIZE

La función JSON_SIZE devuelve el número de bytes de la expresión `SUPER` dada cuando se serializa en una cadena.

### Sintaxis

```
JSON_SIZE(super_expression)
```

### Argumentos

super_expression

Una constante o expresión `SUPER`.

### Tipo de retorno

`INTEGER`

La función JSON_SIZE devuelve un valor `INTEGER` que indica la cantidad de bytes en la cadena de entrada. Este valor es diferente del número de caracteres. Por ejemplo, el carácter UTF-8 #, un punto negro, tiene un tamaño de 3 bytes aunque es 1 carácter.

### Notas de uso

JSON_SIZE (x) es funcionalmente idéntico a OCTET_LENGTH (JSON_SERIALIZE). Sin embargo, tenga en cuenta que JSON_SERIALIZE devuelve un error cuando la expresión SUPER proporcionada supera el límite VARCHAR del sistema cuando se serializa. JSON_SIZE no tiene esta limitación.

### Ejemplos

Para devolver la longitud de un valor SUPER serializado en una cadena, use el siguiente ejemplo.

```
SELECT JSON_SIZE(JSON_PARSE('[10001,10002,"#"]'));

+-----------+
| json_size |
+-----------+
|        19 |
+-----------+
```

Tenga en cuenta que la expresión SUPER proporcionada tiene 17 caracteres, pero # es un carácter de 3 bytes, por lo que JSON_SIZE devuelve 19.

## Función JSON_TYPEOF

La función escalar JSON_TYPEOF devuelve un VARCHAR con valores booleanos, números, cadenas, objetos, matrices o nulos, en función del tipo dinámico del valor SUPER.

### Sintaxis

```
JSON_TYPEOF(super_expression)
```

### Argumentos

super_expression

Una expresión o columna SUPER.

### Tipo de retorno

VARCHAR

### Ejemplos

Para comprobar el tipo de JSON de la matriz `[1,2]` con la función JSON_TYPEOF, utilice el siguiente ejemplo.

```
SELECT JSON_TYPEOF(ARRAY(1,2));

+-------------+
| json_typeof |
+-------------+
| array       |
+-------------+
```

Para comprobar el tipo de JSON del objeto `{"name":"Joe"}` con la función JSON_TYPEOF, utilice el siguiente ejemplo.

```
SELECT JSON_TYPEOF(JSON_PARSE('{"name":"Joe"}'));

+-------------+
| json_typeof |
+-------------+
| object      |
+-------------+
```

## SIZE

Devuelve el tamaño binario en memoria de una constante o expresión de tipo SUPER como un INTEGER.

### Sintaxis

```
SIZE(super_expression)
```

### Argumentos

super_expression

Una constante o expresión de tipo SUPER.

Tipo de retorno

`INTEGER`

Ejemplos

Para usar SIZE para obtener el tamaño en memoria de varias expresiones de tipo SUPER, use el siguiente ejemplo.

```
CREATE TABLE test_super_size(a SUPER);

INSERT INTO test_super_size
VALUES
  (null),
  (TRUE),
  (JSON_PARSE('[0,1,2,3]')),
  (JSON_PARSE('{"a":0,"b":1,"c":2,"d":3}'))
;

SELECT a, SIZE(a)
FROM test_super_size
ORDER BY 2, 1;

+---------------------------+------+
|             a             | size |
+---------------------------+------+
| true                      |    4 |
| NULL                      |    4 |
| [0,1,2,3]                 |   23 |
| {"a":0,"b":1,"c":2,"d":3} |   52 |
+---------------------------+------+
```

## Operadores y funciones VARBYTE

Las funciones y los operadores de Amazon Redshift compatibles con el tipo de datos VARBYTE son los siguientes:

- Operadores VARBYTE
- FROM_HEX
- FROM_VARBYTE
- GETBIT

- TO_HEX
- TO_VARBYTE
- CONCAT
- LEN
- Función LENGTH
- OCTET_LENGTH
- Función SUBSTRING

## Operadores VARBYTE

En la siguiente tabla se muestran los operadores VARBYTE. El operador trabaja con valores binarios del tipo de datos VARBYTE. Si uno o ambos datos de entrada son nulos, el resultado también será nulo.

Operadores admitidos

| Operador | Descripción | Tipo de retorno |
|---|---|---|
| < | Menor que | BOOLEAN |
| <= | Menor que o igual a | BOOLEAN |
| = | Igualdad | BOOLEAN |
| > | Mayor que | BOOLEAN |
| >= | Mayor que o igual a | BOOLEAN |
| != o <> | Distinto de | BOOLEAN |
| \|\| | Concatenación | VARBYTE |
| + | Concatenación | VARBYTE |

| Operador | Descripción | Tipo de retorno |
|---|---|---|
| ~ | NOT bit a bit | VARBYTE |
| & | AND bit a bit | VARBYTE |
| \| | OR bit a bit | VARBYTE |
| # | XOR bit a bit | VARBYTE |

Ejemplos

En los siguientes ejemplos, el valor de `'a'::VARBYTE` es 61, mientras que el valor de `'b'::VARBYTE` es 62. `::` convierte las cadenas en el tipo de datos VARBYTE. Para obtener más información sobre la conversión de los tipos de datos, consulte CAST.

Para comparar si `'a'` es menor que `'b'` con el operador <, utilice el siguiente ejemplo.

```
SELECT 'a'::VARBYTE < 'b'::VARBYTE AS less_than;

+-----------+
| less_than |
+-----------+
| true      |
+-----------+
```

Para comparar si `'a'` es igual que `'b'` con el operador =, utilice el siguiente ejemplo.

```
SELECT 'a'::VARBYTE = 'b'::VARBYTE AS equal;

+-------+
| equal |
+-------+
| false |
+-------+
```

Para concatenar dos valores binarios con el operador | |, utilice el siguiente ejemplo.

```
SELECT 'a'::VARBYTE || 'b'::VARBYTE AS concat;

+--------+
| concat |
+--------+
|   6162 |
+--------+
```

Para concatenar dos valores binarios con el operador +, utilice el siguiente ejemplo.

```
SELECT 'a'::VARBYTE + 'b'::VARBYTE AS concat;

+--------+
| concat |
+--------+
|   6162 |
+--------+
```

Para negar cada bit del valor binario de entrada mediante la función FROM_VARBYTE, utilice el siguiente ejemplo. La cadena `'a'` toma el valor `01100001`. Para obtener más información, consulte FROM_VARBYTE.

```
SELECT FROM_VARBYTE(~'a'::VARBYTE, 'binary');

+--------------+
| from_varbyte |
+--------------+
|     10011110 |
+--------------+
```

Para aplicar el operador & a los dos valores binarios de entrada, use el siguiente ejemplo. La cadena `'a'` se evalúa como `01100001` y `'b'` se evalúa como `01100010`.

```
SELECT FROM_VARBYTE('a'::VARBYTE & 'b'::VARBYTE, 'binary');

+--------------+
| from_varbyte |
+--------------+
|     01100000 |
```

```
+--------------+
```

## Función FROM_HEX

FROM_HEX convierte un hexadecimal en un valor binario.

### Sintaxis

```
FROM_HEX(hex_string)
```

### Argumentos

hex_string

Cadena hexadecimal del tipo de datos VARCHAR o TEXT que hay que convertir. El formato debe ser un valor literal.

### Tipo de retorno

VARBYTE

### Ejemplos

Para convertir la representación hexadecimal de '6162' a un valor binario, use el siguiente ejemplo. El resultado se muestra automáticamente como representación hexadecimal del valor binario.

```
SELECT FROM_HEX('6162');

+----------+
| from_hex |
+----------+
|     6162 |
+----------+
```

## Función FROM_VARBYTE

FROM_VARBYTE convierte un valor binario en una cadena de caracteres en el formato especificado.

### Sintaxis

```
FROM_VARBYTE(binary_value, format)
```

Argumentos

binary_value

Un valor binario del tipo de datos VARBYTE.

formato

Formato de la cadena de caracteres devuelta. Los valores válidos sin distinción entre mayúsculas y minúsculas son `hex`, `binary`, `utf8` (también `utf-8` y `utf_8`) y `base64`.

Tipo de retorno

VARCHAR

Ejemplos

Para convertir el valor binario `'ab'` a hexadecimal, use el siguiente ejemplo.

```
SELECT FROM_VARBYTE('ab', 'hex');

+--------------+
| from_varbyte |
+--------------+
|         6162 |
+--------------+
```

Para devolver la representación binaria de `'4d'`, use el siguiente ejemplo. La representación binaria de `'4d'` corresponde a la cadena de caracteres `01001101`.

```
SELECT FROM_VARBYTE(FROM_HEX('4d'), 'binary');

+--------------+
| from_varbyte |
+--------------+
|     01001101 |
+--------------+
```

## Función GETBIT

GETBIT devuelve el valor de los bits de un valor binario en el índice especificado.

### Sintaxis

```
GETBIT(binary_value, index)
```

### Argumentos

binary_value

Un valor binario del tipo de datos VARBYTE.

índice

Un número de índice del bit en el valor binario que se devuelve. El valor binario es una matriz de bits basada en 0 que se indexa desde el bit situado más a la derecha (bit menos significativo) hasta el bit situado más a la izquierda (bit más significativo).

### Tipo de retorno

INTEGER

### Ejemplos

Para devolver el bit del índice 2 del valor binario `from_hex('4d')`, use el siguiente ejemplo. La representación binaria de `'4d'` es `01001101`.

```
SELECT GETBIT(FROM_HEX('4d'), 2);

+--------+
| getbit |
+--------+
|      1 |
+--------+
```

Para devolver el bit en ocho posiciones de índice correspondiente al valor binario devuelto por `from_hex('4d')`, use el siguiente ejemplo. La representación binaria de `'4d'` es `01001101`.

```
SELECT GETBIT(FROM_HEX('4d'), 7), GETBIT(FROM_HEX('4d'), 6),
  GETBIT(FROM_HEX('4d'), 5), GETBIT(FROM_HEX('4d'), 4),
  GETBIT(FROM_HEX('4d'), 3), GETBIT(FROM_HEX('4d'), 2),
  GETBIT(FROM_HEX('4d'), 1), GETBIT(FROM_HEX('4d'), 0);
```

```
+--------+--------+--------+--------+--------+--------+--------+--------+
| getbit | getbit | getbit | getbit | getbit | getbit | getbit | getbit |
+--------+--------+--------+--------+--------+--------+--------+--------+
|      0 |      1 |      0 |      0 |      1 |      1 |      0 |      1 |
+--------+--------+--------+--------+--------+--------+--------+--------+
```

## Función TO_HEX

TO_HEX convierte un número o valor binario en una representación hexadecimal.

Sintaxis

```
TO_HEX(value)
```

Argumentos

value

Un número o un valor binario (VARBYTE) que hay que convertir.

Tipo de retorno

VARCHAR

Ejemplos

Para convertir un número a su representación hexadecimal, use el siguiente ejemplo.

```
SELECT TO_HEX(2147676847);

+----------+
|  to_hex  |
+----------+
| 8002f2af |
+----------+
```

Para convertir la representación de VARBYTE de `'abc'` en un número hexadecimal, use el siguiente ejemplo.

```
SELECT TO_HEX('abc'::VARBYTE);
```

```
+--------+
| to_hex |
+--------+
| 616263 |
+--------+
```

Para crear una tabla, insertar la representación VARBYTE de `'abc'` en un número hexadecimal y seleccionar la columna con el valor, use el siguiente ejemplo.

```
CREATE TABLE t (vc VARCHAR);
INSERT INTO t SELECT TO_HEX('abc'::VARBYTE);
SELECT vc FROM t;

+--------+
|   vc   |
+--------+
| 616263 |
+--------+
```

Para mostrar cuando se convierte un valor VARBYTE en VARCHAR el formato utilizado es UTF-8, utilice el siguiente ejemplo.

```
CREATE TABLE t (vc VARCHAR);
INSERT INTO t SELECT 'abc'::VARBYTE::VARCHAR;

SELECT vc FROM t;

+-----+
| vc  |
+-----+
| abc |
+-----+
```

## Función TO_VARBYTE

TO_VARBYTE convierte una cadena con un formato especificado en un valor binario.

Sintaxis

```
TO_VARBYTE(string, format)
```

Argumentos

string

Una cadena CHAR o VARCHAR.

formato

El formato de la cadena de entrada. Los valores válidos sin distinción entre mayúsculas y minúsculas son `hex`, `binary`, `utf8` (también `utf-8` y `utf_8`) y `base64`.

Tipo de retorno

VARBYTE

Ejemplos

Para convertir el `6162` hexadecimal en un valor binario, use el siguiente ejemplo. El resultado se muestra automáticamente como representación hexadecimal del valor binario.

```
SELECT TO_VARBYTE('6162', 'hex');

+------------+
| to_varbyte |
+------------+
|       6162 |
+------------+
```

Para devolver la representación binaria de `4d`, use el siguiente ejemplo. La representación binaria de "4d" es `01001101`.

```
SELECT TO_VARBYTE('01001101', 'binary');

+------------+
| to_varbyte |
+------------+
|         4d |
+------------+
```

Para convertir la cadena `'a'` en UTF-8 a un valor binario, use el siguiente ejemplo. El resultado se muestra automáticamente como representación hexadecimal del valor binario.

```
SELECT TO_VARBYTE('a', 'utf8');

+------------+
| to_varbyte |
+------------+
|         61 |
+------------+
```

Para convertir la cadena `'4'` en hexadecimal a un valor binario, use el siguiente ejemplo. Si la longitud de la cadena hexadecimal es un número impar, entonces se le antepone un `0` que formará un número hexadecimal válido.

```
SELECT TO_VARBYTE('4', 'hex');

+------------+
| to_varbyte |
+------------+
|         04 |
+------------+
```

## Funciones de ventana

Con las funciones de ventana, puede crear consultas empresariales analíticas de manera más eficiente. Las funciones de ventana operan en una partición o "ventana" de un conjunto de resultados y devuelven un valor para cada fila de esa ventana. Por el contrario, las funciones que no son de ventana realizan sus cálculos respecto de cada fila en el conjunto de resultados. A diferencia de las funciones de grupo que agregan las filas de resultados, las funciones de ventana retienen todas las filas de la expresión de tabla.

Los valores devueltos se calculan con los valores de los conjuntos de filas en esa ventana. Para cada fila en la tabla, la ventana define un conjunto de filas que se usan para computar atributos adicionales. Una ventana se define utilizando una especificación de ventana (la cláusula OVER) y se basa en tres conceptos principales:

- Particionamiento de ventana, que forma grupos de filas (cláusula PARTITION).
- Ordenación de ventana, que define un orden o una secuencia de filas dentro de cada partición (cláusula ORDER BY).
- Marcos de ventana, que se definen en función de cada fila para restringir aún más el conjunto de filas (especificación ROWS).

Las funciones de ventana son el último conjunto de operaciones realizadas en una consulta, excepto por la cláusula final ORDER BY. Todas las combinaciones y todas las cláusulas WHERE, GROUP BY y HAVING se completan antes de que se procesen las funciones de ventana. Por lo tanto, las funciones de ventana pueden figurar solamente en la lista SELECT o en la cláusula ORDER BY. Se pueden usar distintas funciones de ventana dentro de una única consulta con diferentes cláusulas de marco. También se pueden usar las funciones de ventana en otras expresiones escalares, como CASE.

Las funciones de ventana no se pueden anidar. Por ejemplo, una función agregada SUM puede aparecer dentro de una función de ventana SUM, pero una función de ventana SUM no puede aparecer dentro de otra función de ventana SUM. No se admite lo siguiente porque una función de ventana está anidada en otra función de ventana.

```
SELECT SUM(SUM(selectcol) OVER (PARTITION BY ordercol)) OVER (Partition by ordercol)
 FROM t;
```

## Resumen de la sintaxis de la función de ventana

Las funciones de ventana siguen una sintaxis estándar, que es la que se indica a continuación.

```
function (expression) OVER (
[ PARTITION BY expr_list ]
[ ORDER BY order_list [ frame_clause ] ] )
```

Aquí, function es una de las funciones que se describen en esta sección

La apariencia de expr_list es la siguiente.

```
expression | column_name [, expr_list ]
```

El order_list tiene la siguiente apariencia.

```
expression | column_name [ ASC | DESC ]
[ NULLS FIRST | NULLS LAST ]
[, order_list ]
```

La frame_clause tiene la siguiente apariencia.

```
ROWS
```

```
{ UNBOUNDED PRECEDING | unsigned_value PRECEDING | CURRENT ROW } |

{ BETWEEN
{ UNBOUNDED PRECEDING | unsigned_value { PRECEDING | FOLLOWING } | CURRENT ROW}
AND
{ UNBOUNDED FOLLOWING | unsigned_value { PRECEDING | FOLLOWING } | CURRENT ROW }}
```

Argumentos

función

La función de ventana. Para obtener más información, consulte las descripciones de las funciones individuales.

OVER

La cláusula que define la especificación de ventana. La cláusula OVER es obligatoria para las funciones de ventana y distingue funciones de ventana de otras funciones SQL.

PARTITION BY expr_list

(Opcional) La cláusula PARTITION BY subdivide el conjunto de resultado en particiones, muy similar a la cláusula GROUP BY. Si hay una cláusula de partición, la función se calcula para las filas en cada partición. Si no se especifica una cláusula de partición, una única partición tiene toda la tabla y la función se computa para la tabla completa.

Las funciones de clasificación, DENSE_RANK, NTILE, RANK y ROW_NUMBER, requieren una comparación global de todas las filas en el conjunto de resultados. Cuando se utiliza una cláusula PARTITION BY, el optimizador de consultas puede ejecutar cada agregación en paralelo mediante la distribución de la carga de trabajo entre distintos sectores, según las particiones. Si no hay cláusula PARTITION BY, el paso de agregación se debe ejecutar en serie en un único sector, lo que puede tener consecuencias negativas importantes en el rendimiento, especialmente en el caso de clústeres grandes.

Amazon Redshift no admite literales de cadena en cláusulas PARTITION BY.

ORDER BY order_list

(Opcional) La función de ventana se aplica a las filas dentro de cada partición ordenada, según la especificación de orden en ORDER BY. Esta cláusula ORDER BY es distinta y no guarda relación alguna con las cláusulas ORDER BY de frame_clause. La cláusula ORDER BY se puede usar sin la cláusula PARTITION BY.

Para las funciones de clasificación, la cláusula ORDER BY identifica las medidas para los valores de clasificación. Para las funciones de agregación, las filas particionadas se deben ordenar antes de que la función de agregación se compute para cada marco. Para obtener más información acerca de los tipos de funciones de ventana, consulte Funciones de ventana.

Se requieren identificadores de columnas o expresiones que toman el valor de los identificadores de columnas en la lista de ordenación. No se pueden usar constantes ni expresiones constantes como sustitutos para los nombres de columnas.

Los valores NULLS se tratan como su propio grupo; se ordenan y se clasificación según la opción NULLS FIRST o NULLS LAST. De manera predeterminada, los valores NULL se ordenan y clasificación al final en orden ASC, y se ordenan y se clasifican primero en orden DESC.

Amazon Redshift no admite literales de cadena en cláusulas ORDER BY.

Si se omite la cláusula ORDER BY, el orden de las filas no es determinista.

> **Note**
>
> En cualquier sistema paralelo, como Amazon Redshift, cuando una cláusula ORDER BY no produce una ordenación total y única de los datos, el orden de las filas no es determinístico. Es decir, si la expresión ORDER BY produce valores duplicados (una ordenación parcial), el orden de retorno de esas filas puede variar de una ejecución de Amazon Redshift a la siguiente. A su vez, las funciones de ventana pueden devolver resultados inesperados o inconsistente. Para obtener más información, consulte Ordenación única de datos para funciones de ventana.

column_name

Nombre de una columna que se particionará u ordenará.

ASC | DESC

Opción que define el orden de ordenación para la expresión, de la siguiente manera:

- ASC: ascendente (por ejemplo, de menor a mayor para valores numéricos y de la A a la Z para cadenas con caracteres). Si no se especifica ninguna opción, los datos se ordenan, de manera predeterminada, en orden ascendente.
- DESC: descendente (de mayor a menor para valores numéricos y de la Z a la A para cadenas).

NULLS FIRST | NULLS LAST

Opción que especifica si los valores NULL se deben ordenar en primer lugar, antes de los valores no nulos, o al final, después de los valores no nulos. De manera predeterminada, los valores NULL se ordenan y clasificación al final en orden ASC, y se ordenan y se clasifican primero en orden DESC.

frame_clause

Para funciones de agregación, la cláusula de marco limita el conjunto de filas en una ventana de función al usar ORDER BY. Le permite incluir o excluir conjuntos de filas dentro del resultado ordenado. La cláusula de marco consta de la palabra clave ROWS y de los especificadores correspondientes.

La cláusula de marco no se aplica a las funciones de clasificación. Además, no es necesaria cuando no se utiliza una cláusula ORDER BY en la cláusula OVER para una función de agrupación. Si se utiliza una cláusula ORDER BY para una función de agregación, se necesita una cláusula de marco explícita.

Cuando no se especifica una cláusula ORDER BY, el marco implícito es ilimitado, lo que es equivalente a ROWS BETWEEN UNBOUNDED PRECEDING AND UNBOUNDED FOLLOWING.

ROWS

Esta cláusula define el marco de ventana especificando un desplazamiento físico desde la fila actual .

Esta cláusula especifica las filas en la ventana o partición actual que se combinará con el valor de la fila actual. Utiliza argumentos que especifican la posición de la fila, que puede ser antes o después de la fila actual. El punto de referencia para todos los marcos de ventana es la fila actual. Cada fila se convierte en la fila actual cuando el marco de ventana se desplaza hacia delante en la partición.

El marco puede ser un conjunto simple de filas hasta la fila actual, que se incluye.

```
{UNBOUNDED PRECEDING | offset PRECEDING | CURRENT ROW}
```

O bien, puede ser un conjunto de filas entre dos límites.

```
BETWEEN
{ UNBOUNDED PRECEDING | offset { PRECEDING | FOLLOWING } | CURRENT ROW }
```

```
AND
{ UNBOUNDED FOLLOWING | offset { PRECEDING | FOLLOWING } | CURRENT ROW }
```

UNBOUNDED PRECEDING indica que la ventana comienza en la primera fila de la partición; offset PRECEDING indica que la ventana comienza en un número de filas equivalente al valor de desplazamiento antes de la fila actual. UNBOUNDED PRECEDING es el valor predeterminado.

CURRENT ROW indica que la ventana comienza o finaliza en la fila actual.

UNBOUNDED FOLLOWING indica que la ventana finaliza en la última fila de la partición; offset FOLLOWING indica que la ventana finaliza en un número de filas equivalente al valor de desplazamiento después de la fila actual.

offset identifica un número físico de filas antes o después de la fila actual. En este caso, offset debe ser una constante que se evalúe como un valor numérico positivo. Por ejemplo, 5 FOLLOWING finaliza el marco de cinco filas después de la fila actual.

Cuando no se especifica BETWEEN, el marco se limita implícitamente a la fila actual. Por ejemplo, `ROWS 5 PRECEDING` equivale a `ROWS BETWEEN 5 PRECEDING AND CURRENT ROW`. Además, `ROWS UNBOUNDED FOLLOWING` equivale a `ROWS BETWEEN CURRENT ROW AND UNBOUNDED FOLLOWING`.

> **Note**
>
> No puede especificar un marco en el que el límite de inicio sea mayor que el límite final. Por ejemplo, no puede especificar ninguno de estos marcos.
>
> ```
> between 5 following and 5 preceding
> between current row and 2 preceding
> between 3 following and current row
> ```

## Ordenación única de datos para funciones de ventana

Si una cláusula ORDER BY para una función de ventana no produce una ordenación total y única de los datos, el orden de las filas no es determinístico. Si la expresión ORDER BY produce valores duplicados (una ordenación parcial), el orden de retorno de esas filas puede variar en distintas ejecuciones. En este caso, las funciones de ventana también pueden devolver resultados inesperados o inconsistentes.

Por ejemplo, la siguiente consulta devuelve resultados diferentes con las múltiples ejecuciones. Estos resultados diferentes se producen porque `order by dateid` no genera una ordenación única de los datos para la función de ventana SUM.

```
select dateid, pricepaid,
sum(pricepaid) over(order by dateid rows unbounded preceding) as sumpaid
from sales
group by dateid, pricepaid;

dateid | pricepaid |   sumpaid
--------+-----------+-------------
1827 |   1730.00 |     1730.00
1827 |    708.00 |     2438.00
1827 |    234.00 |     2672.00
...

select dateid, pricepaid,
sum(pricepaid) over(order by dateid rows unbounded preceding) as sumpaid
from sales
group by dateid, pricepaid;

dateid | pricepaid |   sumpaid
--------+-----------+-------------
1827 |    234.00 |      234.00
1827 |    472.00 |      706.00
1827 |    347.00 |     1053.00
...
```

En este caso, agregar una segunda columna ORDER BY a la función de ventana puede solucionar el problema.

```
select dateid, pricepaid,
sum(pricepaid) over(order by dateid, pricepaid rows unbounded preceding) as sumpaid
from sales
group by dateid, pricepaid;

dateid | pricepaid | sumpaid
--------+-----------+---------
1827 |    234.00 |  234.00
1827 |    337.00 |  571.00
1827 |    347.00 |  918.00
...
```

## Funciones compatibles

Amazon Redshift admite dos tipos de funciones de ventana: agrupación y clasificación.

A continuación, se indican las funciones de agregado admitidas:

- Función de ventana AVG
- Función de ventana COUNT
- Función de ventana CUME_DIST
- Función de ventana DENSE_RANK
- Función de ventana FIRST_VALUE
- Función de ventana LAG
- Función de ventana LAST_VALUE
- Función de ventana LEAD
- Función de ventana LISTAGG
- Función de ventana MAX
- Función de ventana MEDIAN
- Función de ventana MIN
- Función de ventana DENSE_NTH
- Función de ventana PERCENTILE_CONT
- Función de ventana PERCENTILE_DISC
- Función de ventana RATIO_TO_REPORT
- Funciones de ventana STDDEV_SAMP y STDDEV_POP (STDDEV_SAMP y STDDEV son sinónimos)
- Función de ventana SUM
- Funciones de ventana VAR_SAMP y VAR_POP (VAR_SAMP y VARIANCE son sinónimos)

A continuación, se indican las funciones de clasificación admitidas:

- Función de ventana DENSE_RANK
- Función de ventana NTILE
- Función de ventana PERCENT_RANK

- Función de ventana RANK
- Función de ventana ROW_NUMBER

## Tabla de muestra para ejemplos de funciones de ventana

Puede encontrar ejemplos específicos de funciones de ventana con la descripción de cada función. Algunos de los ejemplos utilizan una tabla denominada WINSALES, que tiene 11 filas, como se muestra a continuación.

| SALESID | DATEID | SELLERID | BUYERID | QTY | QTY_SHIPPED |
|---|---|---|---|---|---|
| 30001 | 8/2/2003 | 3 | B | 10 | 10 |
| 10001 | 12/24/2003 | 1 | C | 10 | 10 |
| 10005 | 12/24/2003 | 1 | A | 30 | |
| 40001 | 1/9/2004 | 4 | A | 40 | |
| 10006 | 1/18/2004 | 1 | C | 10 | |
| 20001 | 2/12/2004 | 2 | B | 20 | 20 |
| 40005 | 2/12/2004 | 4 | A | 10 | 10 |
| 20002 | 2/16/2004 | 2 | C | 20 | 20 |
| 30003 | 4/18/2004 | 3 | B | 15 | |
| 30004 | 4/18/2004 | 3 | B | 20 | |
| 30007 | 9/7/2004 | 3 | C | 30 | |

El siguiente script crea y completa la tabla de muestra WINSALES.

```
CREATE TABLE winsales(
  salesid int,
  dateid date,
```

```
  sellerid int,
  buyerid char(10),
  qty int,
  qty_shipped int);

INSERT INTO winsales VALUES
  (30001, '8/2/2003', 3, 'b', 10, 10),
  (10001, '12/24/2003', 1, 'c', 10, 10),
  (10005, '12/24/2003', 1, 'a', 30, null),
  (40001, '1/9/2004', 4, 'a', 40, null),
  (10006, '1/18/2004', 1, 'c', 10, null),
  (20001, '2/12/2004', 2, 'b', 20, 20),
  (40005, '2/12/2004', 4, 'a', 10, 10),
  (20002, '2/16/2004', 2, 'c', 20, 20),
  (30003, '4/18/2004', 3, 'b', 15, null),
  (30004, '4/18/2004', 3, 'b', 20, null),
  (30007, '9/7/2004', 3, 'c', 30, null);
```

## Función de ventana AVG

La función de ventana AVG devuelve el promedio (media aritmética) de los valores de la expresión de entrada. La función AVG funciona con valores numéricos e ignora los valores NULL.

### Sintaxis

```
AVG ( [ALL ] expression ) OVER
(
[ PARTITION BY expr_list ]
[ ORDER BY order_list
                        frame_clause ]
)
```

### Argumentos

expression

La columna o expresión de destino sobre la que opera la función.

ALL

Con el argumento ALL, la función retiene todos los valores duplicados de la expresión para el conteo. El valor predeterminado es ALL. DISTINCT no se admite.

OVER

Especifica las cláusulas de ventana para las funciones de agregación. La cláusula OVER distingue funciones de agregación de ventana de las funciones de agregación de conjuntos normales.

PARTITION BY expr_list

Define la ventana para la función AVG en términos de una o más expresiones.

ORDER BY order_list

Ordena las filas dentro de cada partición. Si no se especifica PARTITION BY, ORDER BY utiliza toda la tabla.

frame_clause

Si se utiliza una cláusula ORDER BY para una función de agregación, se necesita una cláusula de marco explícita. La cláusula de marco limita el conjunto de filas en una ventana de función e incluye o excluye conjuntos de filas dentro del resultado ordenado. La cláusula de marco consta de la palabra clave ROWS y de los especificadores correspondientes. Consulte Resumen de la sintaxis de la función de ventana.

### Tipos de datos

Los tipos de argumento compatibles con la función AVG son SMALLINT, INTEGER, BIGINT, NUMERIC, DECIMAL, REAL y DOUBLE PRECISION.

Los tipos de retorno compatibles con la función AVG son los siguientes:

- BIGINT para los argumentos SMALLINT o INTEGER
- NUMERIC para argumentos BIGINT
- DOUBLE PRECISION para argumentos de punto flotante

### Ejemplos

En el siguiente ejemplo, se calcula un promedio móvil de las cantidades vendidas por fecha; ordene los resultados por ID de fecha e ID de ventas:

```
select salesid, dateid, sellerid, qty,
avg(qty) over
(order by dateid, salesid rows unbounded preceding) as avg
```

```
from winsales
order by 2,1;

salesid |   dateid   | sellerid | qty | avg
---------+------------+----------+-----+-----
30001 | 2003-08-02 |        3 |  10 |  10
10001 | 2003-12-24 |        1 |  10 |  10
10005 | 2003-12-24 |        1 |  30 |  16
40001 | 2004-01-09 |        4 |  40 |  22
10006 | 2004-01-18 |        1 |  10 |  20
20001 | 2004-02-12 |        2 |  20 |  20
40005 | 2004-02-12 |        4 |  10 |  18
20002 | 2004-02-16 |        2 |  20 |  18
30003 | 2004-04-18 |        3 |  15 |  18
30004 | 2004-04-18 |        3 |  20 |  18
30007 | 2004-09-07 |        3 |  30 |  19
(11 rows)
```

Para ver una descripción de la tabla WINSALES, consulte Tabla de muestra para ejemplos de funciones de ventana.

## Función de ventana COUNT

La función de ventana COUNT cuenta las filas definidas por la expresión.

La función COUNT tiene dos variaciones. COUNT(*) cuenta todas las filas en la tabla destino, incluya o no valores nulos. COUNT (expresión) computa la cantidad de filas con valores no NULL en una columna o expresión específicas.

### Sintaxis

```
COUNT ( * | [ ALL ] expression) OVER
(
[ PARTITION BY expr_list ]
[ ORDER BY order_list
                        frame_clause ]
)
```

### Argumentos

expression

La columna o expresión de destino sobre la que opera la función.

ALL

Con el argumento ALL, la función retiene todos los valores duplicados de la expresión para el conteo. El valor predeterminado es ALL. DISTINCT no se admite.

OVER

Especifica las cláusulas de ventana para las funciones de agregación. La cláusula OVER distingue funciones de agregación de ventana de las funciones de agregación de conjuntos normales.

PARTITION BY expr_list

Define la ventana para la función COUNT en términos de una o más expresiones.

ORDER BY order_list

Ordena las filas dentro de cada partición. Si no se especifica PARTITION BY, ORDER BY utiliza toda la tabla.

frame_clause

Si se utiliza una cláusula ORDER BY para una función de agregación, se necesita una cláusula de marco explícita. La cláusula de marco limita el conjunto de filas en una ventana de función e incluye o excluye conjuntos de filas dentro del resultado ordenado. La cláusula de marco consta de la palabra clave ROWS y de los especificadores correspondientes. Consulte Resumen de la sintaxis de la función de ventana.

### Tipos de datos

La función COUNT admite todos los tipos de datos de argumentos.

El tipo de retorno compatible con la función COUNT es BIGINT.

### Ejemplos

En el siguiente ejemplo, se muestran los ID de ventas, la cantidad y el conteo de todas las filas desde el comienzo de la ventana de datos:

```
select salesid, qty,
count(*) over (order by salesid rows unbounded preceding) as count
from winsales
```

```
order by salesid;

salesid | qty | count
---------+-----+-----
10001 |  10 |   1
10005 |  30 |   2
10006 |  10 |   3
20001 |  20 |   4
20002 |  20 |   5
30001 |  10 |   6
30003 |  15 |   7
30004 |  20 |   8
30007 |  30 |   9
40001 |  40 |   10
40005 |  10 |   11
(11 rows)
```

Para ver una descripción de la tabla WINSALES, consulte Tabla de muestra para ejemplos de funciones de ventana.

En el siguiente ejemplo, se muestran los ID de ventas, la cantidad y el conteo de las filas no nulas desde el comienzo de la ventana de datos. (En la tabla WINSALES, la columna QTY_SHIPPED tiene algunos valores NULL.)

```
select salesid, qty, qty_shipped,
count(qty_shipped)
over (order by salesid rows unbounded preceding) as count
from winsales
order by salesid;

salesid | qty | qty_shipped | count
---------+-----+-------------+-------
10001 |  10 |          10 |   1
10005 |  30 |             |   1
10006 |  10 |             |   1
20001 |  20 |          20 |   2
20002 |  20 |          20 |   3
30001 |  10 |          10 |   4
30003 |  15 |             |   4
30004 |  20 |             |   4
30007 |  30 |             |   4
40001 |  40 |             |   4
40005 |  10 |          10 |   5
```

```
(11 rows)
```

## Función de ventana CUME_DIST

Calcula la distribución acumulada de un valor dentro de una ventana o partición. Si se asume un orden ascendente, la distribución acumulada se determina utilizando esta fórmula:

```
count of rows with values <= x / count of rows in the window or partition
```

donde x equivale al valor en la fila actual de la columna especificada en la cláusula ORDER BY. El siguiente conjunto de datos ilustra el uso de esta fórmula:

```
Row# Value   Calculation    CUME_DIST
1        2500    (1)/(5)    0.2
2        2600    (2)/(5)    0.4
3        2800    (3)/(5)    0.6
4        2900    (4)/(5)    0.8
5        3100    (5)/(5)    1.0
```

El rango de valor de retorno es > 0 a 1, inclusive.

### Sintaxis

```
CUME_DIST ()
OVER (
[ PARTITION BY partition_expression ]
[ ORDER BY order_list ]
)
```

### Argumentos

OVER

Una cláusula que especifica la partición de ventana. La cláusula OVER no puede tener una especificación de marco de ventana.

PARTITION BY partition_expression

Opcional. Una expresión que establece el rango de registros para cada grupo en la cláusula OVER.

ORDER BY order_list

La expresión sobre la cual se calcula la distribución acumulada. La expresión debe tener un tipo de dato numérico o ser implícitamente convertible a un dato numérico. Si se omite ORDER BY, el valor de retorno es 1 para todas las filas.

Si ORDER BY no produce una ordenación única, el orden de las filas no es determinístico. Para obtener más información, consulte Ordenación única de datos para funciones de ventana.

Tipo de retorno

FLOAT8

Ejemplos

En el siguiente ejemplo, se calcula la distribución acumulada de la cantidad para cada vendedor:

```
select sellerid, qty, cume_dist()
over (partition by sellerid order by qty)
from winsales;

sellerid   qty    cume_dist
------------------------------------------------
1         10.00    0.33
1         10.64    0.67
1         30.37    1
3         10.04    0.25
3         15.15    0.5
3         20.75    0.75
3         30.55    1
2         20.09    0.5
2         20.12    1
4         10.12    0.5
4         40.23    1
```

Para ver una descripción de la tabla WINSALES, consulte Tabla de muestra para ejemplos de funciones de ventana.

## Función de ventana DENSE_RANK

La función de ventana DENSE_RANK determina la clasificación de un valor en un grupo de valores, según la expresión ORDER BY en la cláusula OVER. Si hay una cláusula opcional PARTITION BY,

las clasificaciones de restablecen para cada grupo de filas. Las filas con valores iguales para el criterio de clasificación reciben la misma clasificación. La función DENSE_RANK difiere de RANK en un aspecto: si se vinculan dos o más filas, no hay brecha en la secuencia de valores clasificados. Por ejemplo, si dos filas tienen clasificación 1, la siguiente clasificación es 2.

Puede tener funciones de clasificación con diferentes cláusulas PARTITION BY y ORDER BY en la misma consulta.

Sintaxis

```
DENSE_RANK() OVER
(
[ PARTITION BY expr_list ]
[ ORDER BY order_list ]
)
```

Argumentos

( )

La función no toma argumentos, pero se necesitan los paréntesis vacíos.

OVER

Las cláusulas de ventana para la función DENSE_RANK.

PARTITION BY expr_list

(Opcional) Una o más expresiones que definen la ventana.

ORDER BY order_list

(Opcional) La expresión en que se basan los valores de clasificación. Si no se especifica PARTITION BY, ORDER BY utiliza toda la tabla. Si se omite ORDER BY, el valor de retorno es 1 para todas las filas.

Si ORDER BY no produce una ordenación única, el orden de las filas no es determinístico. Para obtener más información, consulte Ordenación única de datos para funciones de ventana.

Tipo de retorno

`BIGINT`

Ejemplos

En los siguientes ejemplos se utiliza la tabla de ejemplos para las funciones de ventana. Para obtener más información, consulte Tabla de muestra para ejemplos de funciones de ventana.

En el siguiente ejemplo, se ordena la tabla según la cantidad vendida y se asigna a cada fila tanto una clasificación densa como una regular. Los resultados se ordenan después de que se apliquen los resultados de la función de ventana.

```
SELECT salesid, qty,
DENSE_RANK() OVER(ORDER BY qty DESC) AS d_rnk,
RANK() OVER(ORDER BY qty DESC) AS rnk
FROM winsales
ORDER BY 2,1;

+---------+-----+-------+-----+
| salesid | qty | d_rnk | rnk |
+---------+-----+-------+-----+
|   10001 |  10 |     5 |   8 |
|   10006 |  10 |     5 |   8 |
|   30001 |  10 |     5 |   8 |
|   40005 |  10 |     5 |   8 |
|   30003 |  15 |     4 |   7 |
|   20001 |  20 |     3 |   4 |
|   20002 |  20 |     3 |   4 |
|   30004 |  20 |     3 |   4 |
|   10005 |  30 |     2 |   2 |
|   30007 |  30 |     2 |   2 |
|   40001 |  40 |     1 |   1 |
+---------+-----+-------+-----+
```

Tenga en cuenta la diferencia entre las clasificaciones asignadas al mismo conjunto de filas cuando se usan las funciones DENSE_RANK y RANK en simultáneo en la misma consulta.

En el siguiente ejemplo, se divide la tabla según sellerid, se ordena cada partición según la cantidad y se asigna una clasificación densa a cada fila. Los resultados se ordenan después de que se apliquen los resultados de la función de ventana.

```
SELECT salesid, sellerid, qty,
DENSE_RANK() OVER(PARTITION BY sellerid ORDER BY qty DESC) AS d_rnk
FROM winsales
ORDER BY 2,3,1;
```

```
+---------+----------+-----+-------+
| salesid | sellerid | qty | d_rnk |
+---------+----------+-----+-------+
|   10001 |        1 |  10 |     2 |
|   10006 |        1 |  10 |     2 |
|   10005 |        1 |  30 |     1 |
|   20001 |        2 |  20 |     1 |
|   20002 |        2 |  20 |     1 |
|   30001 |        3 |  10 |     4 |
|   30003 |        3 |  15 |     3 |
|   30004 |        3 |  20 |     2 |
|   30007 |        3 |  30 |     1 |
|   40005 |        4 |  10 |     2 |
|   40001 |        4 |  40 |     1 |
+---------+----------+-----+-------+
```

Para utilizar el último ejemplo con éxito, utilice el siguiente comando para insertar una fila en la tabla WINSALES. Esta fila tiene el mismo buyerid, sellerid y qtysold que otra fila. Esto hará que dos filas coincidan en el último ejemplo y, por lo tanto, mostrará la diferencia entre las funciones DENSE_RANK y RANK.

```
INSERT INTO winsales VALUES(30009, '2/2/2003', 3, 'b', 20, NULL);
```

En el siguiente ejemplo, se divide la tabla según buyerid y sellerid, se ordena cada partición según la cantidad y se asigna a cada fila una clasificación densa y regular. Los resultados se ordenan después de que se aplique la función de ventana.

```
SELECT salesid, sellerid, qty, buyerid,
DENSE_RANK() OVER(PARTITION BY buyerid, sellerid ORDER BY qty DESC) AS d_rnk,
RANK() OVER (PARTITION BY buyerid, sellerid ORDER BY qty DESC) AS rnk
FROM winsales
ORDER BY rnk;

+---------+----------+-----+---------+-------+-----+
| salesid | sellerid | qty | buyerid | d_rnk | rnk |
+---------+----------+-----+---------+-------+-----+
|   20001 |        2 |  20 | b       |     1 |   1 |
|   30007 |        3 |  30 | c       |     1 |   1 |
|   10006 |        1 |  10 | c       |     1 |   1 |
|   10005 |        1 |  30 | a       |     1 |   1 |
|   20002 |        2 |  20 | c       |     1 |   1 |
```

```
|   30009 |        3 |  20 | b       |     1 |   1 |
|   40001 |        4 |  40 | a       |     1 |   1 |
|   30004 |        3 |  20 | b       |     1 |   1 |
|   10001 |        1 |  10 | c       |     1 |   1 |
|   40005 |        4 |  10 | a       |     2 |   2 |
|   30003 |        3 |  15 | b       |     2 |   3 |
|   30001 |        3 |  10 | b       |     3 |   4 |
+---------+----------+-----+---------+-------+-----+
```

## Función de ventana FIRST_VALUE

Dado un conjunto ordenado de filas, FIRST_VALUE devuelve el valor de la expresión especificada respecto de la primera fila en el marco de ventana.

Para obtener información sobre cómo seleccionar la última fila del marco, consulte Función de ventana LAST_VALUE.

### Sintaxis

```
FIRST_VALUE( expression )[ IGNORE NULLS | RESPECT NULLS ]
OVER (
[ PARTITION BY expr_list ]
[ ORDER BY order_list frame_clause ]
)
```

### Argumentos

expresión

La columna o expresión de destino sobre la que opera la función.

IGNORE NULLS

Cuando se utiliza esta opción con FIRST_VALUE, la función devuelve el primer valor en el marco que no sea NULL (o NULL si todos los valores son NULL).

RESPECT NULLS

Indica que Amazon Redshift debe incluir valores nulos al momento de determinar qué fila utilizar. De manera predeterminada, se admite RESPECT NULLS si no especifica IGNORE NULLS.

OVER

Introduce las cláusulas de ventana para la función.

PARTITION BY expr_list

Define la ventana para la función en términos de una o más expresiones.

ORDER BY order_list

Ordena las filas dentro de cada partición. Si no se especifica cláusula PARTITION BY, ORDER BY ordena toda la tabla. Si especifica una cláusula ORDER BY, también debe especificar una frame_clause (cláusula_de_marco).

Los resultados de la función FIRST_VALUE dependen del orden de los datos. En los siguientes casos, los resultados son no determinísticos:

- Cuando no se especifica una cláusula ORDER BY y una partición tiene dos valores diferentes para una expresión
- Cuando la expresión toma valores diferentes que corresponden al mismo valor en la lista ORDER BY.

frame_clause

Si se utiliza una cláusula ORDER BY para una función de agregación, se necesita una cláusula de marco explícita. La cláusula de marco limita el conjunto de filas en una ventana de función e incluye o excluye conjuntos de filas en del resultado ordenado. La cláusula de marco consta de la palabra clave ROWS y de los especificadores correspondientes. Consulte Resumen de la sintaxis de la función de ventana.

Tipo de retorno

Estas funciones admiten expresiones que utilizan tipos de datos de Amazon Redshift primitivos. El tipo de retorno es el mismo que el tipo de datos de la expression.

Ejemplos

En los siguientes ejemplos se utiliza la tabla VENUE de los datos de muestra de TICKIT. Para obtener más información, consulte Base de datos de muestra.

El siguiente ejemplo devuelve la capacidad de asientos para cada lugar en la tabla VENUE, con los resultados ordenados por capacidad (de mayor a menor). La función FIRST_VALUE se utiliza para seleccionar el nombre del lugar que corresponda a la primera fila en el marco: en este caso, la fila con la mayor cantidad de asientos. Los resultados se particionan por estado, por lo que cuando cambia el valor VENUESTATE, se selecciona un nuevo primer valor. El marco de ventana está ilimitado de modo que el primer valor se selecciona para cada fila en cada partición.

Para California, `Qualcomm Stadium` tiene la mayor cantidad de asientos (`70561`), por lo que nombre es el primer valor para todas las filas en la partición `CA`.

```
select venuestate, venueseats, venuename,
first_value(venuename)
over(partition by venuestate
order by venueseats desc
rows between unbounded preceding and unbounded following)
from (select * from venue where venueseats >0)
order by venuestate;

venuestate | venueseats |           venuename            |         first_value
-----------+------------+--------------------------------
+----------------------------
CA         |      70561 | Qualcomm Stadium               | Qualcomm Stadium
CA         |      69843 | Monster Park                   | Qualcomm Stadium
CA         |      63026 | McAfee Coliseum                | Qualcomm Stadium
CA         |      56000 | Dodger Stadium                 | Qualcomm Stadium
CA         |      45050 | Angel Stadium of Anaheim       | Qualcomm Stadium
CA         |      42445 | PETCO Park                     | Qualcomm Stadium
CA         |      41503 | AT&T Park                      | Qualcomm Stadium
CA         |      22000 | Shoreline Amphitheatre         | Qualcomm Stadium
CO         |      76125 | INVESCO Field                  | INVESCO Field
CO         |      50445 | Coors Field                    | INVESCO Field
DC         |      41888 | Nationals Park                 | Nationals Park
FL         |      74916 | Dolphin Stadium                | Dolphin Stadium
FL         |      73800 | Jacksonville Municipal Stadium | Dolphin Stadium
FL         |      65647 | Raymond James Stadium          | Dolphin Stadium
FL         |      36048 | Tropicana Field                | Dolphin Stadium
...
```

En el siguiente ejemplo, se muestra el uso de la opción IGNORE NULLS y se utiliza la adición de una nueva fila a la tabla VENUE:

```
insert into venue values(2000,null,'Stanford','CA',90000);
```

Esta nueva fila tiene un valor NULL para la columna VENUENAME. Ahora, se repite la consulta FIRST_VALUE que se mostró antes en esta sección:

```
select venuestate, venueseats, venuename,
first_value(venuename)
over(partition by venuestate
```

```
order by venueseats desc
rows between unbounded preceding and unbounded following)
from (select * from venue where venueseats >0)
order by venuestate;

venuestate | venueseats |         venuename          | first_value
-----------+------------+----------------------------+-------------
CA         |      90000 | NULL                       | NULL
CA         |      70561 | Qualcomm Stadium           | NULL
CA         |      69843 | Monster Park               | NULL
...
```

Debido a que la nueva fila tiene el valor más alto de VENUESEATS (90000) y VENUENAME es NULL, la función FIRST_VALUE devuelve NULL para la partición CA. Para ignorar filas como esta en la evaluación de la función, se agrega la opción IGNORE NULLS al argumento de la función:

```
select venuestate, venueseats, venuename,
first_value(venuename) ignore nulls
over(partition by venuestate
order by venueseats desc
rows between unbounded preceding and unbounded following)
from (select * from venue where venuestate='CA')
order by venuestate;

venuestate | venueseats |         venuename          |   first_value
------------+------------+----------------------------+------------------
CA         |      90000 | NULL                       | Qualcomm Stadium
CA         |      70561 | Qualcomm Stadium           | Qualcomm Stadium
CA         |      69843 | Monster Park               | Qualcomm Stadium
...
```

## Función de ventana LAG

La función de ventana LAG devuelve los valores para una fila en un desplazamiento dado arriba (antes) de la fila actual en la partición.

Sintaxis

```
LAG (value_expr [, offset ])
[ IGNORE NULLS | RESPECT NULLS ]
OVER ( [ PARTITION BY window_partition ] ORDER BY window_ordering )
```

Argumentos

value_expr

La columna o expresión de destino sobre la que opera la función.

desplazamiento

Un parámetro opcional que especifica la cantidad de filas antes de la fila actual para la cual devolver valores. El desplazamiento puede ser un valor entero constante o una expresión que tome un valor entero. Si no especifica un desplazamiento, Amazon Redshift utiliza 1 como valor predeterminado. Un desplazamiento de `0` indica la fila actual.

IGNORE NULLS

Una especificación opcional que indica que Amazon Redshift no debe tener en cuenta los valores nulos al momento de determinar qué fila utilizar. Los valores nulos se incluyen si no se indica IGNORE NULLS.

> **Note**
>
> Puede usar una expresión NVL o COALESCE para reemplazar los valores nulos con otro valor. Para obtener más información, consulte Funciones NVL y COALESCE.

RESPECT NULLS

Indica que Amazon Redshift debe incluir valores nulos al momento de determinar qué fila utilizar. De manera predeterminada, se admite RESPECT NULLS si no especifica IGNORE NULLS.

OVER

Especifica la partición de ventana y el ordenamiento. La cláusula OVER no puede tener una especificación de marco de ventana.

PARTITION BY window_partition

Un argumento opcional que establece el rango de registros para cada grupo en la cláusula OVER.

ORDER BY window_ordering

Ordena las filas dentro de cada partición.

La función de ventana LAG admite expresiones que utilizan cualquiera de los tipos de datos de Amazon Redshift. El tipo de valor devuelto es el mismo que el tipo de la value_expr (expresión_de_valor).

Ejemplos

En el siguiente ejemplo, se muestra la cantidad de tickets vendidos al comprador con un ID de comprador de 3 y la hora en que el comprador 3 compró los tickets. Para comparar cada venta con la venta anterior para el comprador 3, la consulta devuelve la cantidad anterior vendida para cada venta. Debido a que no hay compras antes del 01/16/2008, el primer valor de cantidad vendida anterior es nulo:

```
select buyerid, saletime, qtysold,
lag(qtysold,1) over (order by buyerid, saletime) as prev_qtysold
from sales where buyerid = 3 order by buyerid, saletime;

buyerid |      saletime       | qtysold | prev_qtysold
---------+---------------------+---------+--------------
3 | 2008-01-16 01:06:09 |       1 |
3 | 2008-01-28 02:10:01 |       1 |            1
3 | 2008-03-12 10:39:53 |       1 |            1
3 | 2008-03-13 02:56:07 |       1 |            1
3 | 2008-03-29 08:21:39 |       2 |            1
3 | 2008-04-27 02:39:01 |       1 |            2
3 | 2008-08-16 07:04:37 |       2 |            1
3 | 2008-08-22 11:45:26 |       2 |            2
3 | 2008-09-12 09:11:25 |       1 |            2
3 | 2008-10-01 06:22:37 |       1 |            1
3 | 2008-10-20 01:55:51 |       2 |            1
3 | 2008-10-28 01:30:40 |       1 |            2
(12 rows)
```

## Función de ventana LAST_VALUE

Con un conjunto de filas ordenado, la función LAST_VALUE devuelve el valor de la expresión con respecto a la última fila del marco.

Para obtener información sobre cómo seleccionar la primera fila del marco, consulte Función de ventana FIRST_VALUE.

### Sintaxis

```
LAST_VALUE( expression )[ IGNORE NULLS | RESPECT NULLS ]
OVER (
[ PARTITION BY expr_list ]
[ ORDER BY order_list frame_clause ]
)
```

### Argumentos

expresión

La columna o expresión de destino sobre la que opera la función.

IGNORE NULLS

La función devuelve el último valor en el marco que no sea NULL (o NULL si todos los valores son NULL).

RESPECT NULLS

Indica que Amazon Redshift debe incluir valores nulos al momento de determinar qué fila utilizar. De manera predeterminada, se admite RESPECT NULLS si no especifica IGNORE NULLS.

OVER

Introduce las cláusulas de ventana para la función.

PARTITION BY expr_list

Define la ventana para la función en términos de una o más expresiones.

ORDER BY order_list

Ordena las filas dentro de cada partición. Si no se especifica cláusula PARTITION BY, ORDER BY ordena toda la tabla. Si especifica una cláusula ORDER BY, también debe especificar una frame_clause (cláusula_de_marco).

Los resultados dependen del orden de los datos. En los siguientes casos, los resultados son no determinísticos:

- Cuando no se especifica una cláusula ORDER BY y una partición tiene dos valores diferentes para una expresión
- Cuando la expresión toma valores diferentes que corresponden al mismo valor en la lista ORDER BY.

frame_clause

Si se utiliza una cláusula ORDER BY para una función de agregación, se necesita una cláusula de marco explícita. La cláusula de marco limita el conjunto de filas en una ventana de función e incluye o excluye conjuntos de filas en del resultado ordenado. La cláusula de marco consta de la palabra clave ROWS y de los especificadores correspondientes. Consulte Resumen de la sintaxis de la función de ventana.

Tipo de retorno

Estas funciones admiten expresiones que utilizan tipos de datos de Amazon Redshift primitivos. El tipo de retorno es el mismo que el tipo de datos de la expression.

Ejemplos

En los siguientes ejemplos se utiliza la tabla VENUE de los datos de muestra de TICKIT. Para obtener más información, consulte Base de datos de muestra.

El siguiente ejemplo devuelve la capacidad de asientos para cada lugar en la tabla VENUE, con los resultados ordenados por capacidad (de mayor a menor). La función LAST_VALUE se utiliza para seleccionar el nombre del lugar que corresponda a la última fila en el marco: en este caso, la fila con la menor cantidad de asientos. Los resultados se particionan por estado, por lo que cuando cambia el valor VENUESTATE, se selecciona un nuevo último valor. El marco de la ventana está ilimitado de modo que el último valor se selecciona para cada fila en cada partición.

Para California, se devuelve `Shoreline Amphitheatre` para cada fila en la partición porque tiene la menor cantidad de asientos (`22000`).

```
select venuestate, venueseats, venuename,
last_value(venuename)
over(partition by venuestate
order by venueseats desc
rows between unbounded preceding and unbounded following)
from (select * from venue where venueseats >0)
order by venuestate;

venuestate | venueseats |          venuename            |          last_value
-----------+------------+-------------------------------
+-----------------------------
CA         |      70561 | Qualcomm Stadium              | Shoreline Amphitheatre
```

```
CA         |      69843 | Monster Park                    | Shoreline Amphitheatre
CA         |      63026 | McAfee Coliseum                 | Shoreline Amphitheatre
CA         |      56000 | Dodger Stadium                  | Shoreline Amphitheatre
CA         |      45050 | Angel Stadium of Anaheim        | Shoreline Amphitheatre
CA         |      42445 | PETCO Park                      | Shoreline Amphitheatre
CA         |      41503 | AT&T Park                       | Shoreline Amphitheatre
CA         |      22000 | Shoreline Amphitheatre          | Shoreline Amphitheatre
CO         |      76125 | INVESCO Field                   | Coors Field
CO         |      50445 | Coors Field                     | Coors Field
DC         |      41888 | Nationals Park                  | Nationals Park
FL         |      74916 | Dolphin Stadium                 | Tropicana Field
FL         |      73800 | Jacksonville Municipal Stadium | Tropicana Field
FL         |      65647 | Raymond James Stadium           | Tropicana Field
FL         |      36048 | Tropicana Field                 | Tropicana Field
...
```

## Función de ventana LEAD

La función de ventana LEAD devuelve los valores para una fila en un desplazamiento dado abajo (después) de la fila actual en la partición.

### Sintaxis

```
LEAD (value_expr [, offset ])
[ IGNORE NULLS | RESPECT NULLS ]
OVER ( [ PARTITION BY window_partition ] ORDER BY window_ordering )
```

### Argumentos

value_expr

La columna o expresión de destino sobre la que opera la función.

desplazamiento

Un parámetro opcional que especifica la cantidad de filas debajo de la fila actual para la cual devolver valores. El desplazamiento puede ser un valor entero constante o una expresión que tome un valor entero. Si no especifica un desplazamiento, Amazon Redshift utiliza 1 como valor predeterminado. Un desplazamiento de `0` indica la fila actual.

IGNORE NULLS

Una especificación opcional que indica que Amazon Redshift no debe tener en cuenta los valores nulos al momento de determinar qué fila utilizar. Los valores nulos se incluyen si no se indica IGNORE NULLS.

> **Note**
>
> Puede usar una expresión NVL o COALESCE para reemplazar los valores nulos con otro valor. Para obtener más información, consulte Funciones NVL y COALESCE.

RESPECT NULLS

Indica que Amazon Redshift debe incluir valores nulos al momento de determinar qué fila utilizar. De manera predeterminada, se admite RESPECT NULLS si no especifica IGNORE NULLS.

OVER

Especifica la partición de ventana y el ordenamiento. La cláusula OVER no puede tener una especificación de marco de ventana.

PARTITION BY window_partition

Un argumento opcional que establece el rango de registros para cada grupo en la cláusula OVER.

ORDER BY window_ordering

Ordena las filas dentro de cada partición.

La función de ventana LEAD admite expresiones que utilizan cualquiera de los tipos de datos de Amazon Redshift. El tipo de valor devuelto es el mismo que el tipo de la value_expr (expresión_de_valor).

Ejemplos

En el siguiente ejemplo, se proporciona la comisión para eventos en la tabla SALES para los cuales se vendieron tickets el 1 y el 2 de enero de 2008, y la comisión pagada por la venta de tickets de la venta subsiguiente. El siguiente ejemplo utiliza la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

```
SELECT eventid, commission, saletime, LEAD(commission, 1) over ( ORDER BY saletime ) AS
 next_comm
FROM sales
WHERE saletime BETWEEN '2008-01-09 00:00:00' AND '2008-01-10 12:59:59'
LIMIT 10;

+---------+------------+---------------------+-----------+
| eventid | commission |      saletime       | next_comm |
+---------+------------+---------------------+-----------+
|    1664 |       13.2 | 2008-01-09 01:00:21 |      69.6 |
|     184 |       69.6 | 2008-01-09 01:00:36 |     116.1 |
|    6870 |      116.1 | 2008-01-09 01:02:37 |      11.1 |
|    3718 |       11.1 | 2008-01-09 01:05:19 |     205.5 |
|    6772 |      205.5 | 2008-01-09 01:14:04 |      38.4 |
|    3074 |       38.4 | 2008-01-09 01:26:50 |     209.4 |
|    5254 |      209.4 | 2008-01-09 01:29:16 |      26.4 |
|    3724 |       26.4 | 2008-01-09 01:40:09 |      57.6 |
|    5303 |       57.6 | 2008-01-09 01:40:21 |      51.6 |
|    3678 |       51.6 | 2008-01-09 01:42:54 |      43.8 |
+---------+------------+---------------------+-----------+
```

En el siguiente ejemplo, se muestra la diferencia máxima entre la comisión de los eventos de la tabla SALES y la comisión pagada por la venta de tickets para la posterior venta del mismo evento. Este ejemplo muestra cómo usar LEAD con una cláusula GROUP BY. Como las funciones de ventana no están permitidas en las cláusulas agregadas, en este ejemplo se utiliza una subconsulta. El siguiente ejemplo utiliza la base de datos de ejemplo TICKIT. Para obtener más información, consulte Base de datos de muestra.

```
SELECT eventid, eventname, max(next_comm_diff) as max_commission_difference
FROM
(
    SELECT sales.eventid, eventname, commission - LEAD(commission, 1) over (ORDER BY
 sales.eventid, saletime) AS next_comm_diff
    FROM sales JOIN event ON sales.eventid = event.eventid
)
GROUP BY eventid, eventname
ORDER BY eventid

LIMIT 10

| eventid | eventname                   | max_commission_difference |
+---------+-----------------------------+---------------------------+
```

```
| 1       | Gotterdammerung             | 7.95                       |
| 2       | Boris Godunov               | 227.85                     |
| 3       | Salome                      | 1350.9                     |
| 4       | La Cenerentola (Cinderella) | 790.05                     |
| 5       | Il Trovatore                | 214.05                     |
| 6       | L Elisir d Amore            | 510.9                      |
| 7       | Doctor Atomic               | 180.6                      |
| 9       | The Fly                     | 147                        |
| 10      | Rigoletto                   | 186.6                      |
+---------+-----------------------------+----------------------------+
```

## Función de ventana LISTAGG

Para cada grupo de una consulta, la función de ventana LISTAGG ordena las filas de ese cada grupo según la expresión ORDER BY y luego concatena los valores en una sola cadena.

### Sintaxis

```
LISTAGG( [DISTINCT] expression [, 'delimiter' ] )
[ WITHIN GROUP (ORDER BY order_list) ]
OVER ( [PARTITION BY partition_expression] )
```

### Argumentos

DISTINCT

(Opcional) Una cláusula que elimina los valores duplicados de la expresión especificada antes de la concatenación. Se omiten los espacios posteriores y, por tanto, las cadenas `'a'` y `'a '` se tratan como duplicados. LISTAGG usa el primer valor que se encuentra. Para obtener más información, consulte Importancia de los espacios en blancos anteriores y posteriores.

expresión_de_agregación

Toda expresión válida (como un nombre de columna) que proporcione los valores para la agregación. Se ignoran los valores NULL y las cadenas vacías.

delimiter

(Opcional) La constante de cadena que separará los valores concatenados. El valor predeterminado es NULL.

WITHIN GROUP (ORDER BY order_list)

(Opcional) Una cláusula que especifica el orden de los valores agregados. Determinístico solamente si ORDER BY proporciona un ordenamiento único. La opción predeterminada es agregar todas las filas y devolver un valor único.

OVER

Una cláusula que especifica la partición de ventana. La cláusula OVER no puede contener una especificación de marco de ventana u ordenamiento.

PARTITION BY partition_expression

(Opcional) Establece el rango de registros para cada grupo en la cláusula OVER.

Devuelve

VARCHAR(MAX). Si el conjunto de resultados es mayor que el tamaño máximo VARCHAR (64K - 1 o 65535), LISTAGG devuelve el siguiente error:

```
Invalid operation: Result size exceeds LISTAGG limit
```

Ejemplos

En los siguientes ejemplos, se utiliza la tabla WINSALES. Para ver una descripción de la tabla WINSALES, consulte Tabla de muestra para ejemplos de funciones de ventana.

El siguiente ejemplo devuelve una lista de ID de vendedores ordenados por ID de vendedor.

```
select listagg(sellerid)
within group (order by sellerid)
over() from winsales;

  listagg
------------
 11122333344
...
...
 11122333344
 11122333344
   (11 rows)
```

El siguiente ejemplo devuelve una lista de ID de vendedores para el comprador B, ordenados por fecha.

```
select listagg(sellerid)
within group (order by dateid)
over () as seller
from winsales
where buyerid = 'b' ;

  seller
---------
    3233
    3233
    3233
    3233
```

El siguiente ejemplo devuelve una lista separada por comas de fechas de ventas para el comprador B.

```
select listagg(dateid,',')
within group (order by sellerid desc,salesid asc)
over () as dates
from winsales
where buyerid  = 'b';

             dates
-----------------------------------------
2003-08-02,2004-04-18,2004-04-18,2004-02-12
2003-08-02,2004-04-18,2004-04-18,2004-02-12
2003-08-02,2004-04-18,2004-04-18,2004-02-12
2003-08-02,2004-04-18,2004-04-18,2004-02-12
```

En el siguiente ejemplo, se usa DISTINCT para devolver una lista de fechas de venta distintas para el comprador B.

```
select listagg(distinct dateid,',')
within group (order by sellerid desc,salesid asc)
over () as dates
from winsales
where buyerid  = 'b';

           dates
```

```
-------------------------------
2003-08-02,2004-04-18,2004-02-12
2003-08-02,2004-04-18,2004-02-12
2003-08-02,2004-04-18,2004-02-12
2003-08-02,2004-04-18,2004-02-12
```

El siguiente ejemplo devuelve una lista separada por comas de ID de ventas para cada ID de comprador.

```
select buyerid,
listagg(salesid,',')
within group (order by salesid)
over (partition by buyerid) as sales_id
from winsales
order by buyerid;

+---------+------------------------+
| buyerid |        sales_id        |
+---------+------------------------+
| a       | 10005,40001,40005      |
| a       | 10005,40001,40005      |
| a       | 10005,40001,40005      |
| b       | 20001,30001,30003,30004 |
| b       | 20001,30001,30003,30004 |
| b       | 20001,30001,30003,30004 |
| b       | 20001,30001,30003,30004 |
| c       | 10001,10006,20002,30007 |
| c       | 10001,10006,20002,30007 |
| c       | 10001,10006,20002,30007 |
| c       | 10001,10006,20002,30007 |
+---------+------------------------+
```

## Función de ventana MAX

La función de ventana MAX devuelve el máximo de los valores de la expresión de entrada. La función MAX funciona con valores numéricos e ignora los valores NULL.

### Sintaxis

```
MAX ( [ ALL ] expression ) OVER
(
[ PARTITION BY expr_list ]
```

```
[ ORDER BY order_list frame_clause ]
)
```

### Argumentos

expression

La columna o expresión de destino sobre la que opera la función.

ALL

Con el argumento ALL, la función retiene todos los valores duplicados de la expresión. El valor predeterminado es ALL. DISTINCT no se admite.

OVER

Una cláusula que especifica las cláusulas de ventana para las funciones de agregación. La cláusula OVER distingue funciones de agregación de ventana de las funciones de agregación de conjuntos normales.

PARTITION BY expr_list

Define la ventana para la función MAX en términos de una o más expresiones.

ORDER BY order_list

Ordena las filas dentro de cada partición. Si no se especifica PARTITION BY, ORDER BY utiliza toda la tabla.

frame_clause

Si se utiliza una cláusula ORDER BY para una función de agregación, se necesita una cláusula de marco explícita. La cláusula de marco limita el conjunto de filas en una ventana de función e incluye o excluye conjuntos de filas dentro del resultado ordenado. La cláusula de marco consta de la palabra clave ROWS y de los especificadores correspondientes. Consulte Resumen de la sintaxis de la función de ventana.

### Tipos de datos

Acepta cualquier tipo de datos como entrada. Devuelve el mismo tipo de datos que expresión.

### Ejemplos

En el siguiente ejemplo, se muestran los ID de ventas, la cantidad y la cantidad máxima desde el comienzo de la ventana de datos:

```
select salesid, qty,
max(qty) over (order by salesid rows unbounded preceding) as max
from winsales
order by salesid;

salesid | qty | max
---------+-----+-----
10001 |  10 |  10
10005 |  30 |  30
10006 |  10 |  30
20001 |  20 |  30
20002 |  20 |  30
30001 |  10 |  30
30003 |  15 |  30
30004 |  20 |  30
30007 |  30 |  30
40001 |  40 |  40
40005 |  10 |  40
(11 rows)
```

Para ver una descripción de la tabla WINSALES, consulte Tabla de muestra para ejemplos de funciones de ventana.

En el siguiente ejemplo, se muestran los ID de ventas, la cantidad y la cantidad máxima en un marco restringido:

```
select salesid, qty,
max(qty) over (order by salesid rows between 2 preceding and 1 preceding) as max
from winsales
order by salesid;

salesid | qty | max
---------+-----+-----
10001 |  10 |
10005 |  30 |  10
10006 |  10 |  30
20001 |  20 |  30
20002 |  20 |  20
30001 |  10 |  20
30003 |  15 |  20
30004 |  20 |  15
30007 |  30 |  20
40001 |  40 |  30
```

```
40005 |  10 |  40
(11 rows)
```

## Función de ventana MEDIAN

Calcula el valor mediano para el rango de valores en una ventana o partición. Se ignoran los valores NULL en el rango.

MEDIAN es una función de distribución inversa que asume un modelo de distribución continua.

### Sintaxis

```
MEDIAN ( median_expression )
OVER ( [ PARTITION BY partition_expression ] )
```

### Argumentos

expresión_de_mediana

Una expresión, como un nombre de columna, que proporciona un valor para los cuales determinar la media. La expresión debe tener un tipo de dato numérico o de fecha y hora o ser implícitamente convertible en uno.

OVER

Una cláusula que especifica la partición de ventana. La cláusula OVER no puede contener una especificación de marco de ventana u ordenamiento.

PARTITION BY partition_expression

Opcional. Una expresión que establece el rango de registros para cada grupo en la cláusula OVER.

### Tipos de datos

El tipo de valor devuelto viene determinado por el tipo de datos de expresión_de_mediana.
En la tabla siguiente, se muestra el tipo de valor devuelto para cada tipo de datos de expresión_de_mediana.

| Tipo de entrada | Tipo de retorno |
|---|---|
| INT2, INT4, INT8, NUMERIC, DECIMAL | DECIMAL |

| Tipo de entrada | Tipo de retorno |
| --- | --- |
| FLOAT, DOUBLE | DOBLE |
| DATE | DATE |

Notas de uso

Si el argumento expresión_de_mediana es un tipo de dato DECIMAL definido con la precisión máxima de 38 dígitos, es posible que MEDIAN devuelva un resultado impreciso o un error. Si el valor de retorno de la función MEDIAN supera los 38 dígitos, el resultado se trunca para adaptarse, lo que genera pérdida de precisión. Si, durante la interpolación, un resultado intermedio supera la precisión máxima, se produce un desbordamiento numérico y la función devuelve un error. Para evitar estas condiciones, recomendamos usar un tipo de dato con menor precisión o emitir el argumento median_expression con una precisión menor.

Por ejemplo, la función SUM con un argumento DECIMAL devuelve una precisión predeterminada de 38 dígitos. La escala del resultado es la misma que la escala del argumento. Entonces, por ejemplo, un SUM de una columna DECIMAL(5,2) devuelve un tipo de dato DECIMAL(38,2).

En el siguiente ejemplo, se utiliza una función SUM en el argumento median_expression (expresión_de_mediana) de una función MEDIAN. El tipo de datos de la columna PRICEPAID es DECIMAL (8,2), por lo que la función SUM devuelve un DECIMAL (38,2).

```
select salesid, sum(pricepaid), median(sum(pricepaid))
over() from sales where salesid < 10 group by salesid;
```

Para evitar una posible pérdida de precisión o un error de desbordamiento, convierta el resultado a un tipo de dato DECIMAL con menor precisión, como se muestra en el siguiente ejemplo.

```
select salesid, sum(pricepaid), median(sum(pricepaid)::decimal(30,2))
over() from sales where salesid < 10 group by salesid;
```

Ejemplos

En el siguiente ejemplo, se calcula la cantidad media de ventas para cada vendedor:

```
select sellerid, qty, median(qty)
over (partition by sellerid)
```

```
from winsales
order by sellerid;


sellerid qty median
---------------------------
1  10 10.0
1  10 10.0
1  30 10.0
2  20 20.0
2  20 20.0
3  10 17.5
3  15 17.5
3  20 17.5
3  30 17.5
4  10 25.0
4  40 25.0
```

Para ver una descripción de la tabla WINSALES, consulte Tabla de muestra para ejemplos de funciones de ventana.

## Función de ventana MIN

La función de ventana MIN devuelve el mínimo de los valores de la expresión de entrada. La función MIN funciona con valores numéricos e ignora los valores NULL.

### Sintaxis

```
MIN ( [ ALL ] expression ) OVER
(
[ PARTITION BY expr_list ]
[ ORDER BY order_list frame_clause ]
)
```

### Argumentos

expression

La columna o expresión de destino sobre la que opera la función.

ALL

Con el argumento ALL, la función retiene todos los valores duplicados de la expresión. El valor predeterminado es ALL. DISTINCT no se admite.

OVER

Especifica las cláusulas de ventana para las funciones de agregación. La cláusula OVER distingue funciones de agregación de ventana de las funciones de agregación de conjuntos normales.

PARTITION BY expr_list

Define la ventana para la función MIN en términos de una o más expresiones.

ORDER BY order_list

Ordena las filas dentro de cada partición. Si no se especifica PARTITION BY, ORDER BY utiliza toda la tabla.

frame_clause

Si se utiliza una cláusula ORDER BY para una función de agregación, se necesita una cláusula de marco explícita. La cláusula de marco limita el conjunto de filas en una ventana de función e incluye o excluye conjuntos de filas dentro del resultado ordenado. La cláusula de marco consta de la palabra clave ROWS y de los especificadores correspondientes. Consulte Resumen de la sintaxis de la función de ventana.

Tipos de datos

Acepta cualquier tipo de datos como entrada. Devuelve el mismo tipo de datos que expresión.

Ejemplos

En el siguiente ejemplo, se muestran los ID de ventas, la cantidad y la cantidad mínima desde el comienzo de la ventana de datos:

```
select salesid, qty,
min(qty) over
(order by salesid rows unbounded preceding)
from winsales
order by salesid;

salesid | qty | min
---------+-----+-----
10001 |  10 |  10
10005 |  30 |  10
```

```
10006 |  10 |  10
20001 |  20 |  10
20002 |  20 |  10
30001 |  10 |  10
30003 |  15 |  10
30004 |  20 |  10
30007 |  30 |  10
40001 |  40 |  10
40005 |  10 |  10
(11 rows)
```

Para ver una descripción de la tabla WINSALES, consulte Tabla de muestra para ejemplos de funciones de ventana.

En el siguiente ejemplo, se muestran los ID de ventas, la cantidad y la cantidad mínima en un marco restringido:

```
select salesid, qty,
min(qty) over
(order by salesid rows between 2 preceding and 1 preceding) as min
from winsales
order by salesid;

salesid | qty | min
---------+-----+-----
10001 |  10 |
10005 |  30 |  10
10006 |  10 |  10
20001 |  20 |  10
20002 |  20 |  10
30001 |  10 |  20
30003 |  15 |  10
30004 |  20 |  10
30007 |  30 |  15
40001 |  40 |  20
40005 |  10 |  30
(11 rows)
```

## Función de ventana DENSE_NTH

La función de ventana NTH_VALUE devuelve el valor de la expresión de la fila especificada del marco de ventana relativo a la primera fila de la ventana.

### Sintaxis

```
NTH_VALUE (expr, offset)
[ IGNORE NULLS | RESPECT NULLS ]
OVER
( [ PARTITION BY window_partition ]
[ ORDER BY window_ordering
                        frame_clause ] )
```

### Argumentos

expr

La columna o expresión de destino sobre la que opera la función.

desplazamiento

Determina el número de fila relativo a la primera fila en la ventana para la cual devolver la expresión. El desplazamiento puede ser una constante o una expresión y debe ser un valor entero positivo mayor que 0.

IGNORE NULLS

Una especificación opcional que indica que Amazon Redshift no debe tener en cuenta los valores nulos al momento de determinar qué fila utilizar. Los valores nulos se incluyen si no se indica IGNORE NULLS.

RESPECT NULLS

Indica que Amazon Redshift debe incluir valores nulos al momento de determinar qué fila utilizar. De manera predeterminada, se admite RESPECT NULLS si no especifica IGNORE NULLS.

OVER

Especifica la partición de ventana, el ordenamiento y el marco de ventana.

PARTITION BY window_partition

Establece el rango de registros para cada grupo en la cláusula OVER.

ORDER BY window_ordering

Ordena las filas dentro de cada partición. Si se omite ORDER BY, el marco predeterminado consta de todas las filas en la partición.

frame_clause

Si se utiliza una cláusula ORDER BY para una función de agregación, se necesita una cláusula de marco explícita. La cláusula de marco limita el conjunto de filas en una ventana de función e incluye o excluye conjuntos de filas en del resultado ordenado. La cláusula de marco consta de la palabra clave ROWS y de los especificadores correspondientes. Consulte Resumen de la sintaxis de la función de ventana.

La función de ventana NTH_VALUE admite expresiones que utilizan cualquiera de los tipos de datos de Amazon Redshift. El valor devuelto es del mismo tipo que el valor de expr.

Ejemplos

En el siguiente ejemplo, se muestra la cantidad de asientos en el tercer lugar más grande de California, Florida y Nueva York, comparados con la cantidad de asientos en los demás lugares en esos estados:

```
select venuestate, venuename, venueseats,
nth_value(venueseats, 3)
ignore nulls
over(partition by venuestate order by venueseats desc
rows between unbounded preceding and unbounded following)
as third_most_seats
from (select * from venue where venueseats > 0 and
venuestate in('CA', 'FL', 'NY'))
order by venuestate;

venuestate |           venuename            | venueseats | third_most_seats
-----------+--------------------------------+------------+------------------
CA         | Qualcomm Stadium               |      70561 |            63026
CA         | Monster Park                   |      69843 |            63026
CA         | McAfee Coliseum                |      63026 |            63026
CA         | Dodger Stadium                 |      56000 |            63026
CA         | Angel Stadium of Anaheim       |      45050 |            63026
CA         | PETCO Park                     |      42445 |            63026
CA         | AT&T Park                      |      41503 |            63026
CA         | Shoreline Amphitheatre         |      22000 |            63026
FL         | Dolphin Stadium                |      74916 |            65647
FL         | Jacksonville Municipal Stadium |      73800 |            65647
FL         | Raymond James Stadium          |      65647 |            65647
FL         | Tropicana Field                |      36048 |            65647
NY         | Ralph Wilson Stadium           |      73967 |            20000
```

```
NY         | Yankee Stadium                |      52325 |            20000
NY         | Madison Square Garden         |      20000 |            20000
(15 rows)
```

## Función de ventana NTILE

La función de ventana NTILE divide las filas ordenadas en la partición en la cantidad especificada de grupos clasificados de igual tamaño en la medida que sea posible y devuelve el grupo en el que se inscribe una fila dada.

### Sintaxis

```
NTILE (expr)
OVER (
[ PARTITION BY expression_list ]
[ ORDER BY order_list ]
)
```

### Argumentos

expr

La cantidad de grupos clasificados y debe dar como resultado un valor entero positivo (mayor que 0) para cada partición. El argumento expr no debe admitir valores nulos.

OVER

Una cláusula que especifica la partición y ordenamiento de ventana. La cláusula OVER no puede tener una especificación de marco de ventana.

PARTITION BY window_partition

Opcional. El rango de registros para cada grupo en la cláusula OVER.

ORDER BY window_ordering

Opcional. Una expresión que ordena las filas dentro de cada partición. Si se omite la cláusula ORDER BY, el comportamiento de clasificación es el mismo.

Si ORDER BY no produce un ordenamiento único, el orden de las filas no es determinístico. Para obtener más información, consulte Ordenación única de datos para funciones de ventana.

Tipo de retorno

BIGINT

Ejemplos

En el siguiente ejemplo, se clasifica en cuatro grupos de clasificación el precio pagado por los tickets de Hamlet el 26 de agosto de 2008. El conjunto de resultados es de 17 filas, divididas, casi uniformemente, entre las clasificaciones 1 a 4:

```
select eventname, caldate, pricepaid, ntile(4)
over(order by pricepaid desc) from sales, event, date
where sales.eventid=event.eventid and event.dateid=date.dateid and eventname='Hamlet'
and caldate='2008-08-26'
order by 4;

eventname |  caldate   | pricepaid | ntile
----------+------------+-----------+-------
Hamlet    | 2008-08-26 |   1883.00 |     1
Hamlet    | 2008-08-26 |   1065.00 |     1
Hamlet    | 2008-08-26 |    589.00 |     1
Hamlet    | 2008-08-26 |    530.00 |     1
Hamlet    | 2008-08-26 |    472.00 |     1
Hamlet    | 2008-08-26 |    460.00 |     2
Hamlet    | 2008-08-26 |    355.00 |     2
Hamlet    | 2008-08-26 |    334.00 |     2
Hamlet    | 2008-08-26 |    296.00 |     2
Hamlet    | 2008-08-26 |    230.00 |     3
Hamlet    | 2008-08-26 |    216.00 |     3
Hamlet    | 2008-08-26 |    212.00 |     3
Hamlet    | 2008-08-26 |    106.00 |     3
Hamlet    | 2008-08-26 |    100.00 |     4
Hamlet    | 2008-08-26 |     94.00 |     4
Hamlet    | 2008-08-26 |     53.00 |     4
Hamlet    | 2008-08-26 |     25.00 |     4
(17 rows)
```

## Función de ventana PERCENT_RANK

Calcula la clasificación de porcentaje de una fila dada. La clasificación de porcentaje se determina utilizando la siguiente fórmula:

```
(x - 1) / (the number of rows in the window or partition - 1)
```

donde x es la clasificación de la fila actual. El siguiente conjunto de datos ilustra el uso de esta fórmula:

```
Row# Value Rank Calculation PERCENT_RANK
1 15 1 (1-1)/(7-1) 0.0000
2 20 2 (2-1)/(7-1) 0.1666
3 20 2 (2-1)/(7-1) 0.1666
4 20 2 (2-1)/(7-1) 0.1666
5 30 5 (5-1)/(7-1) 0.6666
6 30 5 (5-1)/(7-1) 0.6666
7 40 7 (7-1)/(7-1) 1.0000
```

El rango de valor de retorno es 0 a 1, inclusive. La primera fila en cualquier conjunto tiene un PERCENT_RANK de 0.

Sintaxis

```
PERCENT_RANK ()
OVER (
[ PARTITION BY partition_expression ]
[ ORDER BY order_list ]
)
```

Argumentos

( )

La función no toma argumentos, pero se necesitan los paréntesis vacíos.

OVER

Una cláusula que especifica la partición de ventana. La cláusula OVER no puede tener una especificación de marco de ventana.

PARTITION BY partition_expression

Opcional. Una expresión que establece el rango de registros para cada grupo en la cláusula OVER.

ORDER BY order_list

Opcional. La expresión sobre la cual se calcula la clasificación de porcentaje. La expresión debe tener un tipo de dato numérico o ser implícitamente convertible a un dato numérico. Si se omite ORDER BY, el valor de retorno es 0 para todas las filas.

Si ORDER BY no produce un ordenamiento único, el orden de las filas no es determinístico. Para obtener más información, consulte Ordenación única de datos para funciones de ventana.

Tipo de retorno

FLOAT8

Ejemplos

En el siguiente ejemplo, se calcula la clasificación de porcentaje de las cantidades de ventas para cada vendedor:

```
select sellerid, qty, percent_rank()
over (partition by sellerid order by qty)
from winsales;

sellerid qty  percent_rank
----------------------------------------
1  10.00  0.0
1  10.64  0.5
1  30.37  1.0
3  10.04  0.0
3  15.15  0.33
3  20.75  0.67
3  30.55  1.0
2  20.09  0.0
2  20.12  1.0
4  10.12  0.0
4  40.23  1.0
```

Para ver una descripción de la tabla WINSALES, consulte Tabla de muestra para ejemplos de funciones de ventana.

## Función de ventana PERCENTILE_CONT

PERCENTILE_CONT es una función de distribución inversa que asume un modelo de distribución continua. Toma un valor percentil y una especificación de ordenación, y devuelve un valor interpolado que se encuadraría dentro de ese valor percentil dado respecto de la especificación de ordenación.

PERCENTILE_CONT computa una interpolación linear entre valores después de ordenarlos. Utilizando el valor percentil `(P)` y la cantidad de filas no nulas `(N)` en el grupo de agregación,

la función computa la cantidad de filas después de ordenar las filas según la especificación de ordenación. Esta cantidad de filas (RN) se computa según la fórmula RN = (1+ (P*(N-1)). El resultado final de la función de agregación se computa por interpolación lineal entre los valores de filas en números de filas CRN = CEILING(RN) y FRN = FLOOR(RN).

El resultado final será el siguiente.

Si (CRN = FRN = RN), entonces el resultado es (value of expression from row at RN)

De lo contrario, el resultado es el siguiente:

(CRN - RN) * (value of expression for row at FRN) + (RN - FRN) * (value of expression for row at CRN).

Puede especificar solamente la cláusula PARTITION en la cláusula OVER. Si se especifica PARTITION, para cada fila, PERCENTILE_CONT devuelve el valor que se encuadraría dentro de ese valor percentil especificado dentro de un conjunto de valores en una partición dada.

Sintaxis

```
PERCENTILE_CONT ( percentile )
WITHIN GROUP (ORDER BY expr)
OVER (  [ PARTITION BY expr_list ]  )
```

Argumentos

percentil

Constante numérica entre 0 y 1. Los valores nulos se ignoran en el cálculo.

WITHIN GROUP ( ORDER BY expr)

Valores numéricos o de fecha/hora específicos para ordenar y calcular el percentil.

OVER

Especifica la partición de ventana. La cláusula OVER no puede contener una especificación de marco de ventana u ordenamiento.

PARTITION BY expr

Argumento opcional que establece el rango de registros para cada grupo en la cláusula OVER.

Devuelve

El tipo de retorno se determina por el tipo de datos de la expresión ORDER BY en la cláusula WITHIN GROUP. En la tabla siguiente, se muestra el tipo de retorno para cada tipo de datos de la expresión ORDER BY.

| Tipo de entrada | Tipo de retorno |
|---|---|
| INT2, INT4, INT8, NUMERIC, DECIMAL | DECIMAL |
| FLOAT, DOUBLE | DOBLE |
| DATE | DATE |
| TIMESTAMP | TIMESTAMP |

Notas de uso

Si la expresión ORDER BY es un tipo de dato DECIMAL definido con la precisión máxima de 38 dígitos, es posible que PERCENTILE_CONT devuelva un resultado impreciso o un error. Si el valor de retorno de la función PERCENTILE_CONT supera los 38 dígitos, el resultado se trunca para adaptarse, lo que genera pérdida de precisión. Si, durante la interpolación, un resultado intermedio supera la precisión máxima, se produce un desbordamiento numérico y la función devuelve un error. Para evitar estas condiciones, recomendamos usar un tipo de dato con menor precisión o emitir la expresión ORDER BY a una precisión menor.

Por ejemplo, la función SUM con un argumento DECIMAL devuelve una precisión predeterminada de 38 dígitos. La escala del resultado es la misma que la escala del argumento. Entonces, por ejemplo, un SUM de una columna DECIMAL(5,2) devuelve un tipo de dato DECIMAL(38,2).

En el siguiente ejemplo, se utiliza una función SUM en la cláusula ORDER BY de una función PERCENTILE_CONT. El tipo de datos de la columna PRICEPAID es DECIMAL (8,2), por lo que la función SUM devuelve un DECIMAL (38,2).

```
select salesid, sum(pricepaid), percentile_cont(0.6)
within group (order by sum(pricepaid) desc) over()
from sales where salesid < 10 group by salesid;
```

Para evitar una posible pérdida de precisión o un error de desbordamiento, convierta el resultado a un tipo de dato DECIMAL con menor precisión, como se muestra en el siguiente ejemplo.

```
select salesid, sum(pricepaid), percentile_cont(0.6)
within group (order by sum(pricepaid)::decimal(30,2) desc) over()
from sales where salesid < 10 group by salesid;
```

### Ejemplos

En los siguientes ejemplos, se utiliza la tabla WINSALES. Para ver una descripción de la tabla WINSALES, consulte Tabla de muestra para ejemplos de funciones de ventana.

```
select sellerid, qty, percentile_cont(0.5)
within group (order by qty)
over() as median from winsales;

 sellerid | qty | median
----------+-----+--------
        1 |  10 |   20.0
        1 |  10 |   20.0
        3 |  10 |   20.0
        4 |  10 |   20.0
        3 |  15 |   20.0
        2 |  20 |   20.0
        3 |  20 |   20.0
        2 |  20 |   20.0
        3 |  30 |   20.0
        1 |  30 |   20.0
        4 |  40 |   20.0
(11 rows)
```

```
select sellerid, qty, percentile_cont(0.5)
within group (order by qty)
over(partition by sellerid) as median from winsales;

 sellerid | qty | median
----------+-----+--------
        2 |  20 |   20.0
        2 |  20 |   20.0
        4 |  10 |   25.0
        4 |  40 |   25.0
        1 |  10 |   10.0
```

```
        1 |  10 |   10.0
        1 |  30 |   10.0
        3 |  10 |   17.5
        3 |  15 |   17.5
        3 |  20 |   17.5
        3 |  30 |   17.5
(11 rows)
```

En el siguiente ejemplo, se calcula PERCENTILE_CONT y PERCENTILE_DISC de las ventas de tickets para vendedores en el estado de Washington.

```
SELECT sellerid, state, sum(qtysold*pricepaid) sales,
percentile_cont(0.6) within group (order by sum(qtysold*pricepaid::decimal(14,2) )
 desc) over(),
percentile_disc(0.6) within group (order by sum(qtysold*pricepaid::decimal(14,2) )
 desc) over()
from sales s, users u
where s.sellerid = u.userid and state = 'WA' and sellerid < 1000
group by sellerid, state;

 sellerid | state |  sales  | percentile_cont | percentile_disc
----------+-------+---------+-----------------+-----------------
      127 | WA    | 6076.00 |         2044.20 |         1531.00
      787 | WA    | 6035.00 |         2044.20 |         1531.00
      381 | WA    | 5881.00 |         2044.20 |         1531.00
      777 | WA    | 2814.00 |         2044.20 |         1531.00
       33 | WA    | 1531.00 |         2044.20 |         1531.00
      800 | WA    | 1476.00 |         2044.20 |         1531.00
        1 | WA    | 1177.00 |         2044.20 |         1531.00
(7 rows)
```

## Función de ventana PERCENTILE_DISC

PERCENTILE_DISC es una función de distribución inversa que asume un modelo de distribución discreta. Toma un valor percentil y una especificación de ordenación, y devuelve un elemento del conjunto dado.

Para un valor percentil dado P, PERCENTILE_DISC ordena los valores de la expresión en la cláusula ORDER BY devuelve el valor con el menor valor de distribución acumulada (con respecto a la misma especificación de ordenamiento) que sea mayor que o igual al P.

Puede especificar solamente la cláusula PARTITION en la cláusula OVER.

### Sintaxis

```
PERCENTILE_DISC ( percentile )
WITHIN GROUP (ORDER BY expr)
OVER (  [ PARTITION BY expr_list ]  )
```

### Argumentos

percentil

Constante numérica entre 0 y 1. Los valores nulos se ignoran en el cálculo.

WITHIN GROUP ( ORDER BY expr)

Valores numéricos o de fecha/hora específicos para ordenar y calcular el percentil.

OVER

Especifica la partición de ventana. La cláusula OVER no puede contener una especificación de marco de ventana u ordenamiento.

PARTITION BY expr

Argumento opcional que establece el rango de registros para cada grupo en la cláusula OVER.

### Devuelve

El mismo tipo de datos que la expresión ORDER BY en la cláusula WITHIN GROUP.

### Ejemplos

En los siguientes ejemplos, se utiliza la tabla WINSALES. Para ver una descripción de la tabla WINSALES, consulte Tabla de muestra para ejemplos de funciones de ventana.

```
SELECT sellerid, qty, PERCENTILE_DISC(0.5)
WITHIN GROUP (ORDER BY qty)
OVER() AS MEDIAN FROM winsales;

+----------+-----+--------+
| sellerid | qty | median |
+----------+-----+--------+
| 3        | 10  | 20     |
| 1        | 10  | 20     |
| 1        | 10  | 20     |
```

```
| 4         | 10  | 20     |
| 3         | 15  | 20     |
| 2         | 20  | 20     |
| 2         | 20  | 20     |
| 3         | 20  | 20     |
| 1         | 30  | 20     |
| 3         | 30  | 20     |
| 4         | 40  | 20     |
+----------+-----+--------+

SELECT sellerid, qty, PERCENTILE_DISC(0.5)
WITHIN GROUP (ORDER BY qty)
OVER(PARTITION BY sellerid) AS MEDIAN FROM winsales;

+----------+-----+--------+
| sellerid | qty | median |
+----------+-----+--------+
| 4         | 10  | 10     |
| 4         | 40  | 10     |
| 3         | 10  | 15     |
| 3         | 15  | 15     |
| 3         | 20  | 15     |
| 3         | 30  | 15     |
| 2         | 20  | 20     |
| 2         | 20  | 20     |
| 1         | 10  | 10     |
| 1         | 10  | 10     |
| 1         | 30  | 10     |
+----------+-----+--------+
```

Para hallar PERCENTILE_DISC(0,25) y PERCENTILE_DISC(0,75) para la cantidad cuando se particiona por el ID del vendedor, utilice los siguientes ejemplos.

```
SELECT sellerid, qty, PERCENTILE_DISC(0.25)
WITHIN GROUP (ORDER BY qty)
OVER(PARTITION BY sellerid) AS quartile1 FROM winsales;

+----------+-----+-----------+
| sellerid | qty | quartile1 |
+----------+-----+-----------+
| 4         | 10  | 10        |
| 4         | 40  | 10        |
| 2         | 20  | 20        |
```

```
| 2         | 20  | 20         |
| 3         | 10  | 10         |
| 3         | 15  | 10         |
| 3         | 20  | 10         |
| 3         | 30  | 10         |
| 1         | 10  | 10         |
| 1         | 10  | 10         |
| 1         | 30  | 10         |
+-----------+-----+------------+

SELECT sellerid, qty, PERCENTILE_DISC(0.75)
WITHIN GROUP (ORDER BY qty)
OVER(PARTITION BY sellerid) AS quartile3 FROM winsales;

+-----------+-----+------------+
| sellerid | qty | quartile3 |
+-----------+-----+------------+
| 3         | 10  | 20         |
| 3         | 15  | 20         |
| 3         | 20  | 20         |
| 3         | 30  | 20         |
| 4         | 10  | 40         |
| 4         | 40  | 40         |
| 2         | 20  | 20         |
| 2         | 20  | 20         |
| 1         | 10  | 30         |
| 1         | 10  | 30         |
| 1         | 30  | 30         |
+-----------+-----+------------+
```

## Función de ventana RANK

La función de ventana RANK determina la clasificación de un valor en un grupo de valores, según la expresión ORDER BY en la cláusula OVER. Si hay una cláusula opcional PARTITION BY, las clasificaciones de restablecen para cada grupo de filas. Las filas con valores iguales para el criterio de clasificación reciben la misma clasificación. Amazon Redshift agrega el número de filas vinculadas a la clasificación vinculada para calcular la siguiente clasificación; por tanto, es posible que las clasificaciones no tengan números consecutivos. Por ejemplo, si dos filas tienen clasificación 1, la siguiente clasificación es 3.

RANK difiere de Función de ventana DENSE_RANK en un aspecto: para DENSE_RANK, si se vinculan dos o más filas, no hay brecha en la secuencia de valores clasificados. Por ejemplo, si dos filas tienen clasificación 1, la siguiente clasificación es 2.

Puede tener funciones de clasificación con diferentes cláusulas PARTITION BY y ORDER BY en la misma consulta.

Sintaxis

```
RANK () OVER
(
[ PARTITION BY expr_list ]
[ ORDER BY order_list ]
)
```

Argumentos

( )

La función no toma argumentos, pero se necesitan los paréntesis vacíos.

OVER

Las cláusulas de ventana para la función RANK.

PARTITION BY expr_list

Opcional. Una o más expresiones que definen la ventana.

ORDER BY order_list

Opcional. Define las columnas en que se basan los valores de clasificación. Si no se especifica PARTITION BY, ORDER BY utiliza toda la tabla. Si se omite ORDER BY, el valor de retorno es 1 para todas las filas.

Si ORDER BY no produce un ordenamiento único, el orden de las filas no es determinístico. Para obtener más información, consulte Ordenación única de datos para funciones de ventana.

Tipo de retorno

INTEGER

## Ejemplos

En el siguiente ejemplo, se ordena la tabla por la cantidad vendida (orden ascendiente predeterminado) y se asigna una clasificación a cada fila. Un valor de 1 es la mejor clasificación. Los resultados se ordenan después de que se apliquen los resultados de la función de ventana:

```
select salesid, qty,
rank() over (order by qty) as rnk
from winsales
order by 2,1;

salesid | qty | rnk
--------+-----+-----
10001 |  10 |  1
10006 |  10 |  1
30001 |  10 |  1
40005 |  10 |  1
30003 |  15 |  5
20001 |  20 |  6
20002 |  20 |  6
30004 |  20 |  6
10005 |  30 |  9
30007 |  30 |  9
40001 |  40 |  11
(11 rows)
```

Tenga en cuenta que la cláusula ORDER BY externa en este ejemplo incluye las columnas 2 y 1 para asegurarse de que Amazon Redshift devuelva los resultados ordenados de manera uniforme cada vez que se ejecute esta consulta. Por ejemplo, las filas con ID de ventas 10001 y 10006 tienen valores QTY y RNK idénticos. Ordenar el resultado final por columna 1 garantiza que la fila 10001 siempre esté antes que la 10006. Para ver una descripción de la tabla WINSALES, consulte Tabla de muestra para ejemplos de funciones de ventana.

En el siguiente ejemplo, la ordenación se invierte para la función de ventana (`order by qty desc`). Ahora, el valor más alto de clasificación se aplica al valor QTY más alto.

```
select salesid, qty,
rank() over (order by qty desc) as rank
from winsales
order by 2,1;
```

```
 salesid | qty | rank
---------+-----+-----
   10001 |  10 |   8
   10006 |  10 |   8
   30001 |  10 |   8
   40005 |  10 |   8
   30003 |  15 |   7
   20001 |  20 |   4
   20002 |  20 |   4
   30004 |  20 |   4
   10005 |  30 |   2
   30007 |  30 |   2
   40001 |  40 |   1
(11 rows)
```

Para ver una descripción de la tabla WINSALES, consulte Tabla de muestra para ejemplos de funciones de ventana.

En el siguiente ejemplo, se divide la tabla según SELLERID, se ordena cada partición según la cantidad (en orden descendiente) y se asigna una clasificación a cada fila. Los resultados se ordenan después de que se apliquen los resultados de la función de ventana.

```
select salesid, sellerid, qty, rank() over
(partition by sellerid
order by qty desc) as rank
from winsales
order by 2,3,1;

salesid | sellerid | qty | rank
--------+----------+-----+-----
  10001 |        1 |  10 |  2
  10006 |        1 |  10 |  2
  10005 |        1 |  30 |  1
  20001 |        2 |  20 |  1
  20002 |        2 |  20 |  1
  30001 |        3 |  10 |  4
  30003 |        3 |  15 |  3
  30004 |        3 |  20 |  2
  30007 |        3 |  30 |  1
  40005 |        4 |  10 |  2
  40001 |        4 |  40 |  1
(11 rows)
```

## Función de ventana RATIO_TO_REPORT

Calcula la relación de un valor con la suma de los valores en una ventana o partición. La relación de un valor de informe se determina utilizando la fórmula:

`value of` ratio_expression `argument for the current row / sum of` ratio_expression `argument for the window or partition`

El siguiente conjunto de datos ilustra el uso de esta fórmula:

```
Row# Value Calculation RATIO_TO_REPORT
1 2500 (2500)/(13900) 0.1798
2 2600 (2600)/(13900) 0.1870
3 2800 (2800)/(13900) 0.2014
4 2900 (2900)/(13900) 0.2086
5 3100 (3100)/(13900) 0.2230
```

El rango de valor de retorno es 0 a 1, inclusive. Si ratio_expression es NULL, el valor de devolución es NULL. Si un valor en partition_expression es único, la función devolverá 1 para ese valor.

Sintaxis

```
RATIO_TO_REPORT ( ratio_expression )
OVER ( [ PARTITION BY partition_expression ] )
```

Argumentos

ratio_expression

Una expresión, como un nombre de columna, que proporciona el valor para el cual determinar la relación. La expresión debe tener un tipo de dato numérico o ser implícitamente convertible a un dato numérico.

No se puede usar ninguna otra función analítica en ratio_expression.

OVER

Una cláusula que especifica la partición de ventana. La cláusula OVER no puede contener una especificación de marco de ventana u ordenamiento.

PARTITION BY partition_expression

Opcional. Una expresión que establece el rango de registros para cada grupo en la cláusula OVER.

Tipo de retorno

FLOAT8

Ejemplos

En los siguientes ejemplos, se utiliza la tabla WINSALES. Para obtener información sobre cómo crear esta tabla WINSALES, consulte Tabla de muestra para ejemplos de funciones de ventana.

En el siguiente ejemplo, se calcula la relación entre el valor registrado de cada fila de la cantidad de un vendedor y el total de todas las cantidades del vendedor.

```
select sellerid, qty, ratio_to_report(qty)
over()
from winsales
order by sellerid;

sellerid  qty    ratio_to_report
---------------------------------------
1         30     0.13953488372093023
1         10     0.046511627906976744
1         10     0.046511627906976744
2         20     0.09302325581395349
2         20     0.09302325581395349
3         30     0.13953488372093023
3         20     0.09302325581395349
3         15     0.06976744186046512
3         10     0.046511627906976744
4         10     0.046511627906976744
4         40     0.18604651162790697
```

En el siguiente ejemplo, se calculan las relaciones de porcentaje de las cantidades de ventas para vendedor por partición.

```
select sellerid, qty, ratio_to_report(qty)
over(partition by sellerid)
```

```
from winsales;

sellerid   qty    ratio_to_report
----------------------------------------
2          20     0.5
2          20     0.5
4          40     0.8
4          10     0.2
1          10     0.2
1          30     0.6
1          10     0.2
3          10     0.13333333333333333
3          15     0.2
3          20     0.26666666666666666
3          30     0.4
```

## Función de ventana ROW_NUMBER

Asigna un número ordinal de la fila actual en un grupo de filas, contando desde 1, según la expresión ORDER BY en la cláusula OVER. Si hay una cláusula opcional PARTITION BY, los números ordinales se restablecen para cada grupo de filas. Las filas con valores iguales para las expresiones ORDER BY reciben los diferentes números de fila de manera no determinística.

### Sintaxis

```
ROW_NUMBER() OVER(
  [ PARTITION BY expr_list ]
  [ ORDER BY order_list ]
)
```

### Argumentos

( )

La función no toma argumentos, pero se necesitan los paréntesis vacíos.

OVER

La cláusula de la función de ventana para la función ROW_NUMBER.

PARTITION BY expr_list

Opcional. Una o varias expresiones de columna que dividen los resultados en conjuntos de filas.

ORDER BY order_list

Opcional. Una o más expresiones de columna que definen el orden de las filas de un conjunto. Si no se especifica PARTITION BY, ORDER BY utiliza toda la tabla.

Si ORDER BY no produce una ordenación única o se omite, el orden de las filas no es determinístico. Para obtener más información, consulte Ordenación única de datos para funciones de ventana.

Tipo de retorno

BIGINT

Ejemplos

En los siguientes ejemplos se usa la tabla `WINSALES`. Para ver una descripción de la tabla `WINSALES`, consulte Tabla de muestra para ejemplos de funciones de ventana.

En el siguiente ejemplo se ordena la tabla por QTY (en orden ascendente) y, a continuación, se asigna un número de fila a cada fila. Los resultados se ordenan después de que se apliquen los resultados de la función de ventana.

```
SELECT salesid, sellerid, qty,
ROW_NUMBER() OVER(
   ORDER BY qty ASC) AS row
FROM winsales
ORDER BY 4,1;

salesid   sellerid   qty   row
---------+----------+-----+-----
   30001 |        3 |  10 |   1
   10001 |        1 |  10 |   2
   10006 |        1 |  10 |   3
   40005 |        4 |  10 |   4
   30003 |        3 |  15 |   5
   20001 |        2 |  20 |   6
   20002 |        2 |  20 |   7
   30004 |        3 |  20 |   8
   10005 |        1 |  30 |   9
   30007 |        3 |  30 |  10
   40001 |        4 |  40 |  11
```

En el siguiente ejemplo, se particiona la tabla según SELLERID y se ordena cada partición según QTY (en orden ascendiente); luego, se asigna un número a cada fila. Los resultados se ordenan después de que se apliquen los resultados de la función de ventana.

```
SELECT salesid, sellerid, qty,
ROW_NUMBER() OVER(
  PARTITION BY sellerid
  ORDER BY qty ASC) AS row_by_seller
FROM winsales
ORDER BY 2,4;

 salesid | sellerid | qty | row_by_seller
---------+----------+-----+-----
   10001 |        1 |  10 |   1
   10006 |        1 |  10 |   2
   10005 |        1 |  30 |   3
   20001 |        2 |  20 |   1
   20002 |        2 |  20 |   2
   30001 |        3 |  10 |   1
   30003 |        3 |  15 |   2
   30004 |        3 |  20 |   3
   30007 |        3 |  30 |   4
   40005 |        4 |  10 |   1
   40001 |        4 |  40 |   2
```

En el siguiente ejemplo se muestran los resultados cuando no se utilizan las cláusulas opcionales.

```
SELECT salesid, sellerid, qty, ROW_NUMBER() OVER() AS row
FROM winsales
ORDER BY 4,1;

salesid   sellerid   qty   row
---------+----------+-----+-----
   30001 |        3 |  10 |   1
   10001 |        1 |  10 |   2
   10005 |        1 |  30 |   3
   40001 |        4 |  40 |   4
   10006 |        1 |  10 |   5
   20001 |        2 |  20 |   6
   40005 |        4 |  10 |   7
   20002 |        2 |  20 |   8
   30003 |        3 |  15 |   9
   30004 |        3 |  20 |  10
```

```
  30007 |        3 |  30 |  11
```

## Funciones de ventana STDDEV_SAMP y STDDEV_POP

Las funciones de ventana STDDEV_SAMP y STDDEV_POP devuelven la muestra y la desviación estándar de población de un conjunto de valores numéricos (entero, decimal o de punto flotante). Véase también Funciones STDDEV_SAMP y STDDEV_POP.

STDDEV_SAMP y STDDEV son sinónimos para la misma función.

### Sintaxis

```
STDDEV_SAMP | STDDEV | STDDEV_POP
( [ ALL ] expression ) OVER
(
[ PARTITION BY expr_list ]
[ ORDER BY order_list
                        frame_clause ]
)
```

### Argumentos

expression

La columna o expresión de destino sobre la que opera la función.

ALL

Con el argumento ALL, la función retiene todos los valores duplicados de la expresión. El valor predeterminado es ALL. DISTINCT no se admite.

OVER

Especifica las cláusulas de ventana para las funciones de agregación. La cláusula OVER distingue funciones de agregación de ventana de las funciones de agregación de conjuntos normales.

PARTITION BY expr_list

Define la ventana para la función en términos de una o más expresiones.

ORDER BY order_list

Ordena las filas dentro de cada partición. Si no se especifica PARTITION BY, ORDER BY utiliza toda la tabla.

frame_clause

Si se utiliza una cláusula ORDER BY para una función de agregación, se necesita una cláusula de marco explícita. La cláusula de marco limita el conjunto de filas en una ventana de función e incluye o excluye conjuntos de filas dentro del resultado ordenado. La cláusula de marco consta de la palabra clave ROWS y de los especificadores correspondientes. Consulte Resumen de la sintaxis de la función de ventana.

Tipos de datos

Los tipos de argumento compatibles con las funciones STDDEV son SMALLINT, INTEGER, BIGINT, NUMERIC, DECIMAL, REAL y DOUBLE PRECISION.

Independientemente del tipo de datos de la expresión, el tipo de retorno de una función STDDEV es un número de doble precisión.

Ejemplos

En el siguiente ejemplo, se muestra cómo usar las funciones STDDEV_POP y VAR_POP como funciones de ventana. La consulta computa la varianza de población y la desviación estándar de población para valores PRICEPAID en la tabla SALES.

```
select salesid, dateid, pricepaid,
round(stddev_pop(pricepaid) over
(order by dateid, salesid rows unbounded preceding)) as stddevpop,
round(var_pop(pricepaid) over
(order by dateid, salesid rows unbounded preceding)) as varpop
from sales
order by 2,1;

salesid | dateid | pricepaid | stddevpop | varpop
--------+--------+-----------+-----------+---------
  33095 |   1827 |    234.00 |         0 |       0
  65082 |   1827 |    472.00 |       119 |   14161
  88268 |   1827 |    836.00 |       248 |   61283
  97197 |   1827 |    708.00 |       230 |   53019
 110328 |   1827 |    347.00 |       223 |   49845
 110917 |   1827 |    337.00 |       215 |   46159
 150314 |   1827 |    688.00 |       211 |   44414
 157751 |   1827 |   1730.00 |       447 |  199679
 165890 |   1827 |   4192.00 |      1185 | 1403323
```

```
...
```

La desviación estándar de muestra y las funciones de varianza se pueden usar de la misma manera.

## Función de ventana SUM

La función de ventana SUM devuelve la suma de la columna de entrada o valores de la expresión. La función SUM funciona con valores numéricos e ignora los valores NULL.

Sintaxis

```
SUM ( [ ALL ] expression ) OVER
(
[ PARTITION BY expr_list ]
[ ORDER BY order_list
                        frame_clause ]
)
```

Argumentos

expression

La columna o expresión de destino sobre la que opera la función.

ALL

Con el argumento ALL, la función retiene todos los valores duplicados de la expresión. El valor predeterminado es ALL. DISTINCT no se admite.

OVER

Especifica las cláusulas de ventana para las funciones de agregación. La cláusula OVER distingue funciones de agregación de ventana de las funciones de agregación de conjuntos normales.

PARTITION BY expr_list

Define la ventana para la función SUM en términos de una o más expresiones.

ORDER BY order_list

Ordena las filas dentro de cada partición. Si no se especifica PARTITION BY, ORDER BY utiliza toda la tabla.

frame_clause

Si se utiliza una cláusula ORDER BY para una función de agregación, se necesita una cláusula de marco explícita. La cláusula de marco limita el conjunto de filas en una ventana de función e incluye o excluye conjuntos de filas dentro del resultado ordenado. La cláusula de marco consta de la palabra clave ROWS y de los especificadores correspondientes. Consulte Resumen de la sintaxis de la función de ventana.

### Tipos de datos

Los tipos de argumento compatibles con la función SUM son SMALLINT, INTEGER, BIGINT, NUMERIC, DECIMAL, REAL y DOUBLE PRECISION.

Los tipos de retorno compatibles con la función SUM son los siguientes:

- BIGINT para los argumentos SMALLINT o INTEGER
- NUMERIC para argumentos BIGINT
- DOUBLE PRECISION para argumentos de punto flotante

### Ejemplos

En el siguiente ejemplo, se crea una suma acumulativa (continua) de cantidades de ventas ordenadas por fecha e ID de ventas:

```
select salesid, dateid, sellerid, qty,
sum(qty) over (order by dateid, salesid rows unbounded preceding) as sum
from winsales
order by 2,1;

salesid |   dateid   | sellerid | qty | sum
---------+------------+----------+-----+-----
30001 | 2003-08-02 |        3 |  10 |  10
10001 | 2003-12-24 |        1 |  10 |  20
10005 | 2003-12-24 |        1 |  30 |  50
40001 | 2004-01-09 |        4 |  40 |  90
10006 | 2004-01-18 |        1 |  10 | 100
20001 | 2004-02-12 |        2 |  20 | 120
40005 | 2004-02-12 |        4 |  10 | 130
20002 | 2004-02-16 |        2 |  20 | 150
30003 | 2004-04-18 |        3 |  15 | 165
30004 | 2004-04-18 |        3 |  20 | 185
```

```
30007 | 2004-09-07 |        3 |  30 | 215
(11 rows)
```

Para ver una descripción de la tabla WINSALES, consulte Tabla de muestra para ejemplos de funciones de ventana.

En el siguiente ejemplo, se crea una suma acumulativa (continua) de cantidades de ventas por fecha, se dividen los resultados por ID de vendedor y se ordenan los resultados por fecha e ID de ventas dentro de la partición:

```
select salesid, dateid, sellerid, qty,
sum(qty) over (partition by sellerid
order by dateid, salesid rows unbounded preceding) as sum
from winsales
order by 2,1;

salesid |   dateid   | sellerid | qty | sum
---------+------------+----------+-----+-----
30001 | 2003-08-02 |        3 |  10 |  10
10001 | 2003-12-24 |        1 |  10 |  10
10005 | 2003-12-24 |        1 |  30 |  40
40001 | 2004-01-09 |        4 |  40 |  40
10006 | 2004-01-18 |        1 |  10 |  50
20001 | 2004-02-12 |        2 |  20 |  20
40005 | 2004-02-12 |        4 |  10 |  50
20002 | 2004-02-16 |        2 |  20 |  40
30003 | 2004-04-18 |        3 |  15 |  25
30004 | 2004-04-18 |        3 |  20 |  45
30007 | 2004-09-07 |        3 |  30 |  75
(11 rows)
```

En el siguiente ejemplo, se enumeran en forma secuencial todas las filas del conjunto de resultados, ordenadas por las columnas SELLERID y SALESID:

```
select salesid, sellerid, qty,
sum(1) over (order by sellerid, salesid rows unbounded preceding) as rownum
from winsales
order by 2,1;

salesid | sellerid |  qty | rownum
--------+----------+------+--------
10001 |        1 |   10 |     1
```

```
10005 |        1 |   30 |     2
10006 |        1 |   10 |     3
20001 |        2 |   20 |     4
20002 |        2 |   20 |     5
30001 |        3 |   10 |     6
30003 |        3 |   15 |     7
30004 |        3 |   20 |     8
30007 |        3 |   30 |     9
40001 |        4 |   40 |    10
40005 |        4 |   10 |    11
(11 rows)
```

Para ver una descripción de la tabla WINSALES, consulte Tabla de muestra para ejemplos de funciones de ventana.

En el siguiente ejemplo, se enumeran en forma secuencial todas las filas en el conjunto de resultados, se dividen los resultados por SELLERID y se ordenan los resultados por SELLERID y SALESID dentro de la partición:

```
select salesid, sellerid, qty,
sum(1) over (partition by sellerid
order by sellerid, salesid rows unbounded preceding) as rownum
from winsales
order by 2,1;

salesid | sellerid | qty | rownum
---------+----------+-----+--------
10001 |        1 |  10 |      1
10005 |        1 |  30 |      2
10006 |        1 |  10 |      3
20001 |        2 |  20 |      1
20002 |        2 |  20 |      2
30001 |        3 |  10 |      1
30003 |        3 |  15 |      2
30004 |        3 |  20 |      3
30007 |        3 |  30 |      4
40001 |        4 |  40 |      1
40005 |        4 |  10 |      2
(11 rows)
```

## Funciones de ventana VAR_SAMP y VAR_POP

Las funciones de ventana VAR_SAMP y VAR_POP devuelven la muestra y la varianza de población de un conjunto de valores numéricos (entero, decimal o de punto flotante). Véase también Funciones VAR_SAMP y VAR_POP.

VAR_SAMP y VARIANCE son sinónimos para la misma función.

### Sintaxis

```
VAR_SAMP | VARIANCE | VAR_POP
( [ ALL ] expression ) OVER
(
[ PARTITION BY expr_list ]
[ ORDER BY order_list
                        frame_clause ]
)
```

### Argumentos

expression

La columna o expresión de destino sobre la que opera la función.

ALL

Con el argumento ALL, la función retiene todos los valores duplicados de la expresión. El valor predeterminado es ALL. DISTINCT no se admite.

OVER

Especifica las cláusulas de ventana para las funciones de agregación. La cláusula OVER distingue funciones de agregación de ventana de las funciones de agregación de conjuntos normales.

PARTITION BY expr_list

Define la ventana para la función en términos de una o más expresiones.

ORDER BY order_list

Ordena las filas dentro de cada partición. Si no se especifica PARTITION BY, ORDER BY utiliza toda la tabla.

frame_clause

Si se utiliza una cláusula ORDER BY para una función de agregación, se necesita una cláusula de marco explícita. La cláusula de marco limita el conjunto de filas en una ventana de función e incluye o excluye conjuntos de filas dentro del resultado ordenado. La cláusula de marco consta de la palabra clave ROWS y de los especificadores correspondientes. Consulte Resumen de la sintaxis de la función de ventana.

Tipos de datos

Los tipos de argumento compatibles con las funciones VARIANCE son SMALLINT, INTEGER, BIGINT, NUMERIC, DECIMAL, REAL y DOUBLE PRECISION.

Independientemente del tipo de datos de la expresión, el tipo de retorno de una función VARIANCE es un número de doble precisión.

## Funciones de administración del sistema

Temas


- CHANGE_QUERY_PRIORITY
- CHANGE_SESSION_PRIORITY
- CHANGE_USER_PRIORITY
- CURRENT_SETTING
- PG_CANCEL_BACKEND
- PG_TERMINATE_BACKEND
- REBOOT_CLUSTER
- SET_CONFIG


Amazon Redshift admite distintas funciones de administración del sistema.

### CHANGE_QUERY_PRIORITY

CHANGE_QUERY_PRIORITY permite a los superusuarios modificar la prioridad de una consulta que se está ejecutando o esperando en la administración de carga de trabajo (WLM).

Esta función permite a los superusuarios cambiar de inmediato la prioridad de cualquier consulta en el sistema. Solo se puede ejecutar una consulta, usuario o sesión con la prioridad `CRITICAL`.

### Sintaxis

```
CHANGE_QUERY_PRIORITY(query_id, priority)
```

### Argumentos

query_id

El identificador de consulta de la consulta cuya prioridad se cambia. Requiere un valor `INTEGER`.

priority

La nueva prioridad que asignar a la cola. Este argumento debe ser una cadena con el valor `CRITICAL`, `HIGHEST`, `HIGH`, `NORMAL`, `LOW` o `LOWEST`.

### Tipo de retorno

Ninguno

### Ejemplos

Para mostrar la columna `query_priority` en la tabla del sistema STV_WLM_QUERY_STATE, use el siguiente ejemplo.

```
SELECT query, service_class, query_priority, state
FROM stv_wlm_query_state WHERE service_class = 101;

+-------+---------------+----------------+---------+
| query | service_class | query_priority |  state  |
+-------+---------------+----------------+---------+
|  1076 |           101 | Lowest         | Running |
|  1075 |           101 | Lowest         | Running |
+-------+---------------+----------------+---------+
```

Para mostrar los resultados de un superusuario que ejecuta la función `change_query_priority` para cambiar la prioridad a `CRITICAL`, use el siguiente ejemplo.

```
SELECT CHANGE_QUERY_PRIORITY(1076, 'Critical');

+-----------------------------------------------------------------------------+
|                           change_query_priority                             |
```

```
+-------------------------------------------------------------------------------+
| Succeeded to change query priority. Priority changed from Lowest to Critical. |
+-------------------------------------------------------------------------------+
```

## CHANGE_SESSION_PRIORITY

CHANGE_SESSION_PRIORITY permite a los superusuarios cambiar de inmediato la prioridad de cualquier sesión en el sistema. Solo se puede ejecutar una sesión, usuario o consulta con la prioridad `CRITICAL`.

### Sintaxis

```
CHANGE_SESSION_PRIORITY(pid, priority)
```

### Argumentos

pid

El identificador de proceso de la sesión cuya prioridad se cambia. El valor `-1` hace referencia a la sesión actual. Requiere un valor `INTEGER`.

priority

La nueva prioridad que asignar a la sesión. Este argumento debe ser una cadena con el valor `CRITICAL`, `HIGHEST`, `HIGH`, `NORMAL`, `LOW` o `LOWEST`.

### Tipo de retorno

Ninguno

### Ejemplos

Para devolver el identificador de proceso del proceso del servidor que administra la sesión actual, use el siguiente ejemplo.

```
SELECT pg_backend_pid();

+----------------+
| pg_backend_pid |
+----------------+
|          30311 |
```

```
+----------------+
```

En este ejemplo, la prioridad se cambia a LOWEST para la sesión actual.

```
SELECT CHANGE_SESSION_PRIORITY(30311, 'Lowest');

+-------------------------------------------------------------------------------------
+
|                               change_session_priority
 |
+-------------------------------------------------------------------------------------
+
| Succeeded to change session priority. Changed session (pid:30311) priority to lowest.
 |
+-------------------------------------------------------------------------------------
+
```

En este ejemplo, la prioridad se cambia a HIGH para la sesión actual.

```
SELECT CHANGE_SESSION_PRIORITY(-1, 'High');

+-------------------------------------------------------------------------------------
+
|                                  change_session_priority
           |
+-------------------------------------------------------------------------------------
+
| Succeeded to change session priority. Changed session (pid:30311) priority from
 lowest to high. |
+-------------------------------------------------------------------------------------
+
```

Para crear un procedimiento almacenado que cambie la prioridad de una sesión, use el siguiente ejemplo. El permiso para ejecutar este procedimiento almacenado se concede al usuario de la base de datos test_user.

```
CREATE OR REPLACE PROCEDURE sp_priority_low(pid IN int, result OUT varchar)
AS $$
BEGIN
  SELECT CHANGE_SESSION_PRIORITY(pid, 'low') into result;
END;
$$ LANGUAGE plpgsql
```

```
SECURITY DEFINER;
GRANT EXECUTE ON PROCEDURE sp_priority_low(int) TO test_user;
```

A continuación, el usuario de la base de datos denominado `test_user` llama al procedimiento.

```
CALL sp_priority_low(pg_backend_pid());

+-------------------------------------------------------+
|                        result                         |
+-------------------------------------------------------+
| Success. Change session (pid:13155) priority to low. |
+-------------------------------------------------------+
```

## CHANGE_USER_PRIORITY

CHANGE_USER_PRIORITY permite a los superusuarios modificar la prioridad de todas las consultas emitidas por un usuario que se están ejecutando o esperando en la administración de carga de trabajo (WLM). Solo se puede ejecutar un usuario, sesión o consulta con la prioridad `CRITICAL`.

Sintaxis

```
CHANGE_USER_PRIORITY(user_name, priority)
```

Argumentos

user_name

El nombre del usuario de base de datos cuya prioridad de consulta se cambia.

priority

La nueva prioridad que asignar a todas las consultas emitidas por `user_name`. Este argumento debe ser una cadena con el valor `CRITICAL`, `HIGHEST`, `HIGH`, `NORMAL`, `LOW`, `LOWEST` o `RESET`. Solo los superusuarios pueden cambiar la prioridad a `CRITICAL`. El cambio de la prioridad a `RESET` elimina la configuración de prioridad para `user_name`.

Tipo de retorno

Ninguno

### Ejemplos

Para cambiar la prioridad para el usuario `analysis_user` a `LOWEST`, use el siguiente ejemplo.

```
SELECT CHANGE_USER_PRIORITY('analysis_user', 'lowest');

+-------------------------------------------------------------------------------------+
|                               change_user_priority                                  |
+-------------------------------------------------------------------------------------+
| Succeeded to change user priority. Changed user (analysis_user) priority to lowest. |
+-------------------------------------------------------------------------------------+
```

Para cambiar la prioridad a `LOW`, utilice el siguiente ejemplo.

```
SELECT CHANGE_USER_PRIORITY('analysis_user', 'low');

+-------------------------------------------------------------------------------------------------------------
+
|                                    change_user_priority
        |
+-------------------------------------------------------------------------------------------------------------
+
| Succeeded to change user priority. Changed user (analysis_user) priority from Lowest
 to low. |
+-------------------------------------------------------------------------------------------------------------
+
```

Para restablecer la prioridad, utilice el siguiente ejemplo.

```
SELECT CHANGE_USER_PRIORITY('analysis_user', 'reset');

+-------------------------------------------------------+
|                change_user_priority                   |
+-------------------------------------------------------+
| Succeeded to reset priority for user (analysis_user). |
+-------------------------------------------------------+
```

## CURRENT_SETTING

CURRENT_SETTING devuelve el valor actual del parámetro de configuración especificado.

La función equivale al comando SHOW.

Sintaxis

```
current_setting('parameter')
```

En la instrucción siguiente se devuelve el valor actual de la variable de contexto de sesión especificada.

```
current_setting('variable_name')
current_setting('variable_name'[, error_if_undefined])
```

Argumentos

parameter

Valor de parámetro por mostrar. Para obtener una lista de parámetros de configuración, consulte Referencia de la configuración.

variable_name

El nombre de la variable para mostrar. Debe ser una constante de cadena para las variables de contexto de sesión.

error_if_undefined

(Opcional) Un valor booleano opcional que especifica el comportamiento si el nombre de la variable no existe. Cuando error_if_undefined se establece en `TRUE`, de forma predeterminada, Amazon Redshift emite un error. Cuando error_if_undefined se establece en `FALSE`, Amazon Redshift devuelve `NULL`. Amazon Redshift admite el parámetro error_if_undefined solo para variables de contexto de sesión. No se puede utilizar cuando la entrada es un parámetro de configuración.

Tipo de retorno

Devuelve una cadena `CHAR` o `VARCHAR`.

Ejemplos

Para devolver la configuración actual para el parámetro `query_group`, use el siguiente ejemplo.

```
SELECT CURRENT_SETTING('query_group');
```

```
+-----------------+
| current_setting |
+-----------------+
| unset           |
+-----------------+
```

Para devolver la configuración actual para la variable `app_context.user_id`, use el siguiente ejemplo.

```
SELECT CURRENT_SETTING('app_context.user_id', FALSE);
```

## PG_CANCEL_BACKEND

Cancela una consulta. PG_CANCEL_BACKEND es funcionalmente equivalente al comando CANCEL. Puede cancelar consultas que su usuario esté ejecutando actualmente. Los superusuarios pueden cancelar cualquier consulta.

Sintaxis

```
pg_cancel_backend( pid )
```

Argumentos

pid

El ID de proceso (PID) de la consulta que se cancelará. No puede cancelar una consulta al especificar un ID de consulta; debe especificar el ID de proceso de la consulta. Requiere un valor `INTEGER`.

Tipo de retorno

Ninguno

Notas de uso

Si las consultas de varias sesiones contienen bloqueos en la misma tabla, puede usar la función PG_TERMINATE_BACKEND para terminar una de las sesiones, lo que exige que las transacciones en ejecución de la sesión terminada liberen todos los bloqueos y reviertan la transacción. Consulte la tabla de catálogo PG_LOCKS para ver los bloqueos actuales. Si no puede cancelar una consulta

porque está en un bloque de transacción (BEGIN ... END), puede terminar la sesión en la que se ejecuta la consulta al usar la función PG_TERMINATE_BACKEND.

Ejemplos

Para cancelar una consulta que se está ejecutando actualmente, primero recupere el ID de proceso para la consulta que desea cancelar. Para determinar los ID de proceso para todas las consultas en ejecución actualmente, ejecute el siguiente comando.

```
SELECT pid, TRIM(starttime) AS start,
duration, TRIM(user_name) AS user,
SUBSTRING(query,1,40) AS querytxt
FROM stv_recents
WHERE status = 'Running';

+-----+----------------------+----------+--------+----------------------------+
| pid |       starttime      | duration |  user  |          querytxt          |
+-----+----------------------+----------+--------+----------------------------+
| 802 | 2013-10-14 09:19:03.55 |      132 | dwuser | select venuename from venue |
| 834 | 2013-10-14 08:33:49.47 |  1250414 | dwuser | select * from listing;      |
| 964 | 2013-10-14 08:30:43.29 |   326179 | dwuser | select sellerid from sales  |
+-----+----------------------+----------+--------+----------------------------+
```

Para cancelar la consulta con el ID de proceso 802, utilice el siguiente ejemplo.

```
SELECT PG_CANCEL_BACKEND(802);
```

## PG_TERMINATE_BACKEND

Termina una sesión. Puede terminar una sesión propiedad de su usuario. Un superusuario puede terminar cualquier sesión.

Sintaxis

```
pg_terminate_backend( pid )
```

Argumentos

pid

El ID de proceso de la sesión que se terminará. Requiere un valor `INTEGER`.

Tipo de retorno

Ninguno

Notas de uso

Si está cerca de alcanzar el límite para las conexiones simultáneas, utilice PG_TERMINATE_BACKEND para terminar las sesiones inactivas y liberar conexiones. Para obtener más información, consulte Límites de Amazon Redshift.

Si las consultas de distintas sesiones tienen bloqueos en la misma tabla, puede usar la función PG_TERMINATE_BACKEND para terminar una de las sesiones, lo que fuerza que las transacciones en ejecución de la sesión terminada liberen todos los bloqueos y reviertan la transacción. Consulte la tabla de catálogo PG_LOCKS para ver los bloqueos actuales.

Si una consulta no está en un bloque de transacción (BEGIN ... END), puede cancelar la consulta utilizando el comando CANCEL o la función PG_CANCEL_BACKEND.

Ejemplos

Para consultar la tabla SVV_TRANSACTIONS para ver todos los bloqueos vigentes para las transacciones actuales, use el siguiente ejemplo.

```
SELECT * FROM svv_transactions;

+----------+--------+-------+------+--------------------+-----------------
+----------------------+----------+---------+
| txn_owner | txn_db |  xid  | pid  |      txn_start      |    lock_mode    |
 lockable_object_type | relation | granted |
+----------+--------+-------+------+--------------------+-----------------
+----------------------+----------+---------+
| rsuser    | dev    | 96178 | 8585 | 2017-04-12 20:13:07 | AccessShareLock | relation
           |    51940 | true    |
| rsuser    | dev    | 96178 | 8585 | 2017-04-12 20:13:07 | AccessShareLock | relation
           |    52000 | true    |
| rsuser    | dev    | 96178 | 8585 | 2017-04-12 20:13:07 | AccessShareLock | relation
           |   108623 | true    |
| rsuser    | dev    | 96178 | 8585 | 2017-04-12 20:13:07 | ExclusiveLock   |
 transactionid        |          | true    |
+----------+--------+-------+------+--------------------+-----------------
+----------------------+----------+---------+
```

Para finalizar la sesión que contiene los bloqueos, utilice el siguiente ejemplo.

```
SELECT PG_TERMINATE_BACKEND(8585);
```

## REBOOT_CLUSTER

Reinicie el clúster de Amazon Redshift sin cerrar sus conexiones. Debe ser un superusuario de base de datos para ejecutar este comando.

Una vez que finaliza este reinicio flexible, el clúster de Amazon Redshift devuelve un error a la aplicación del usuario y requiere que la aplicación del usuario vuelva a enviar toda transacción o consulta interrumpida por dicho proceso.

Sintaxis

```
SELECT REBOOT_CLUSTER();
```

## SET_CONFIG

Establece un parámetro de configuración para una configuración nueva.

La función equivale al comando SET en SQL.

Sintaxis

```
SET_CONFIG('parameter', 'new_value' , is_local)
```

La siguiente instrucción establece una variable de contexto de sesión en una nueva configuración.

```
set_config('variable_name', 'new_value' , is_local)
```

Argumentos

parameter

Parámetro que se establecerá.

variable_name

El nombre de la variable que establecer.

new_value

Valor nuevo del parámetro.

is_local

Si es verdadero, el valor de parámetro se aplica únicamente a la transacción actual. Los valores válidos son `true` o `1` y `false` o `0`.

Tipo de retorno

Devuelve una cadena `CHAR` o `VARCHAR`.

Ejemplos

Para establecer el valor del parámetro `query_group` para `test` solo para la transacción actual, use el siguiente ejemplo.

```
SELECT SET_CONFIG('query_group', 'test', true);

+------------+
| set_config |
+------------+
| test       |
+------------+
```

Para establecer las variables de contexto de sesión, use el siguiente ejemplo.

```
SELECT SET_CONFIG('app.username', 'cuddy', FALSE);
```

## Funciones de información del sistema

Amazon Redshift admite distintas funciones de información del sistema.

Temas

- CURRENT_AWS_ACCOUNT
- CURRENT_DATABASE
- CURRENT_NAMESPACE
- CURRENT_SCHEMA
- CURRENT_SCHEMAS
- CURRENT_SESSION_ARN
- CURRENT_USER

- CURRENT_USER_ID
- DEFAULT_IAM_ROLE
- GET_MOUNTED_ROLE
- HAS_ASSUMEROLE_PRIVILEGE
- HAS_DATABASE_PRIVILEGE
- HAS_SCHEMA_PRIVILEGE
- HAS_TABLE_PRIVILEGE
- LAST_USER_QUERY_ID
- PG_BACKEND_PID
- PG_GET_COLS
- PG_GET_GRANTEE_BY_IAM_ROLE
- PG_GET_IAM_ROLE_BY_USER
- PG_GET_LATE_BINDING_VIEW_COLS
- PG_GET_SESSION_ROLES
- PG_LAST_COPY_COUNT
- PG_LAST_COPY_ID
- PG_LAST_UNLOAD_ID
- PG_LAST_QUERY_ID
- PG_LAST_UNLOAD_COUNT
- Función SLICE_NUM
- USER
- ROLE_IS_MEMBER_OF
- USER_IS_MEMBER_OF
- VERSION

## CURRENT_AWS_ACCOUNT

Devuelve la cuenta de AWS asociada al clúster de Amazon Redshift que envió una consulta.

Sintaxis

```
current_aws_account
```

Tipo de retorno

Devuelve un número entero.

Ejemplo

La siguiente consulta devuelve el nombre de la base de datos actual.

```
select user, current_aws_account;
current_user | current_account
-------------+---------------
dwuser       | 987654321

(1 row)
```

## CURRENT_DATABASE

Devuelve el nombre de la base de datos a la cual está conectado actualmente.

Sintaxis

```
current_database()
```

Tipo de retorno

Devuelve una cadena CHAR o VARCHAR.

Ejemplo

La siguiente consulta devuelve el nombre de la base de datos actual.

```
select current_database();

current_database
------------------
tickit
(1 row)
```

## CURRENT_NAMESPACE

Devuelve el espacio de nombres del clúster de Amazon Redshift actual. El espacio de nombres del clúster de Amazon Redshift es el ID único del clúster de Amazon Redshift.

Sintaxis

```
current_namespace
```

Tipo de retorno

Devuelve una cadena CHAR o VARCHAR.

Ejemplo

La siguiente consulta devuelve el nombre del espacio de nombres actual.

```
select user, current_namespace;
current_user | current_namespace
-------------+--------------------------------------
dwuser       | 86b5169f-01dc-4a6f-9fbb-e2e24359e9a8

(1 row)
```

## CURRENT_SCHEMA

Devuelve el nombre del esquema al frente de la ruta de búsqueda. Este esquema se usará para cualquier tabla u otros objetos con nombre que se creen sin especificar un esquema de destino.

Sintaxis

> **Note**
>
> Esta es una función del nodo principal. Esta función devuelve un error si hace referencia a una tabla creada por usuarios, a una tabla de sistema STL o STV, o a una vista de sistema SVV o SVL.

```
current_schema()
```

Tipo de retorno

CURRENT_SCHEMA devuelve una cadena VARCHAR o CHAR.

Ejemplos

La siguiente consulta devuelve el esquema actual:

```
select current_schema();

current_schema
----------------
public
(1 row)
```

## CURRENT_SCHEMAS

Devuelve una matriz de nombres de cualquier esquema en la ruta de búsqueda actual. La ruta de búsqueda actual se define en el parámetro search_path.

Sintaxis

> **Note**
>
> Esta es una función del nodo principal. Esta función devuelve un error si hace referencia a una tabla creada por usuarios, a una tabla de sistema STL o STV, o a una vista de sistema SVV o SVL.

```
current_schemas(include_implicit)
```

Argumento

include_implicit

Si es verdadero, especifica que la ruta de búsqueda debe incluir todos los esquemas de sistema incluidos implícitamente. Los valores válidos son `true` y `false`. Por lo general, si es `true`, este parámetro devuelve el esquema `pg_catalog`, además del esquema actual.

Tipo de retorno

Devuelve una cadena CHAR o VARCHAR.

Ejemplos

El siguiente ejemplo devuelve los nombres de los esquemas en la ruta de búsqueda actual, sin incluir los esquemas de sistema incluidos implícitamente:

```
select current_schemas(false);

current_schemas
-----------------
{public}
(1 row)
```

El siguiente ejemplo devuelve los nombres de los esquemas en la ruta de búsqueda actual, incluidos los esquemas de sistema incluidos implícitamente:

```
select current_schemas(true);

current_schemas
--------------------
{pg_catalog,public}
(1 row)
```

## CURRENT_SESSION_ARN

Devuelve el ARN del usuario global autorizado actualmente. Los usuarios globales se presentan con la misma identidad en todas las cuentas, clústeres y grupos de trabajo sin servidor de Redshift. Los usuarios globales inician sesión a través de IAM Identity Center o mediante la autenticación de sesión basada en IAM. Los usuarios del lago de datos son usuarios de AWS globales.

Esta función se suele utilizar en el contexto del uso de vistas multidialectales de AWS Glue. Para obtener más información sobre la administración de identidades con IAM Identity Center y Redshift, consulte Conectar Redshift con IAM Identity Center para ofrecer a los usuarios una experiencia de inicio de sesión único. Para obtener más información sobre las vistas de Glue multidialectales, consulte Creación de vistas en el catálogo de datos de AWS Glue.

### Sintaxis

```
current_session_arn()
```

### Tipo de retorno

Devuelve una cadena VARCHAR del usuario autenticado de forma global o un valor nulo.

### Notas de uso

No se admiten los usuarios locales y el resultado es una respuesta nula.

Ejemplo

La siguiente consulta devuelve el nombre del ARN de la sesión actual:

```
SELECT current_session_arn();

current_session_arn
--------------
arn:aws:iam::123456789012:user/user
(1 row)
```

## CURRENT_USER

Devuelve el nombre de usuario del usuario "vigente" actual de la base de datos, aplicable a los permisos de comprobación. Por lo general, este nombre de usuario será el mismo usuario de la sesión; no obstante, ocasionalmente los superusuarios pueden cambiarlo.

> **Note**
>
> No utilice paréntesis a la derecha al ejecutar CURRENT_USER.

Sintaxis

```
current_user
```

Tipo de retorno

CURRENT_USER devuelve un tipo de datos NAME y se puede convertir en una cadena CHAR o VARCHAR.

Notas de uso

Si se creó un procedimiento almacenado mediante la opción SECURITY DEFINER del comando CREATE_PROCEDURE, al invocar la función CURRENT_USER desde el procedimiento almacenado, Amazon Redshift devuelve el nombre de usuario del propietario del procedimiento almacenado.

Ejemplo

La siguiente consulta devuelve el nombre de usuario de la base de datos actual:

```
select current_user;

current_user
--------------
dwuser
(1 row)
```

## CURRENT_USER_ID

Devuelve un identificador único para el usuario de Amazon Redshift que ha abierto la sesión actual.

Sintaxis

```
CURRENT_USER_ID
```

Tipo de retorno

La función CURRENT_USER_ID devuelve un valor entero.

Ejemplos

El siguiente ejemplo devuelve el nombre de usuario y el ID de usuario actual para esta sesión:

```
select user, current_user_id;

 current_user | current_user_id
--------------+-----------------
   dwuser     |               1
(1 row)
```

## DEFAULT_IAM_ROLE

Devuelve el rol de IAM predeterminado que está actualmente asociado con el clúster de Amazon Redshift. La función no devuelve ningún valor si no hay ningún rol de IAM asociado de manera predeterminada.

Sintaxis

```
select default_iam_role();
```

Tipo de retorno

Devuelve una cadena VARCHAR.

Ejemplo

El siguiente ejemplo devuelve el rol de IAM predeterminado que está actualmente asociado con el clúster de Amazon Redshift especificado.

```
select default_iam_role();
              default_iam_role
--------------------------------------------
 arn:aws:iam::123456789012:role/myRedshiftRole
(1 row)
```

## GET_MOUNTED_ROLE

Cuando se invoca como parte de una vista multidialectal de AWS Glue, permite devolver el rol de IAM utilizado para montar el esquema o la base de datos de Lake Formation. Multidialectal significa que SQL es compatible con varios motores de consulta, como Amazon EMR y Redshift. Para obtener más información sobre las vistas de Glue multidialectales, consulte Creación de vistas en el catálogo de datos de AWS Glue.

Sintaxis

```
get_mounted_role()
```

Tipo de retorno

Devuelve una cadena VARCHAR o un valor nulo.

Notas de uso

Esta función devuelve un valor nulo para cualquier caso de uso fuera de una vista externa de Lake Formation.

Ejemplo

La siguiente consulta devuelve la identidad para montar el recurso de Lake Formation.

```
CREATE EXTERNAL PROTECTED VIEW external_schema.remote_view AS
```

```
SELECT mycol, get_mounted_role() FROM external_schema.remote_table;

mycol | get_mounted_role
--------------------------
1       arn:aws:iam::123456789012:role/salesrole
(1 row)
```

## HAS_ASSUMEROLE_PRIVILEGE

Devuelve el valor booleano `true` (`t`) si el usuario en cuestión tiene el rol de IAM especificado con el privilegio necesario para ejecutar el comando especificado. La función devuelve `false` (`f`) si el usuario no tiene el rol de IAM especificado con el privilegio necesario para ejecutar el comando especificado. Para obtener más información acerca de los privilegios, consulte GRANT.

### Sintaxis

```
has_assumerole_privilege( [ user, ] iam_role_arn, cmd_type)
```

### Argumentos

usuario

El nombre del usuario para verificar los privilegios del rol de IAM. El valor predeterminado es comprobar el usuario actual. Los superusuarios y los usuarios pueden usar esta función. No obstante, los usuarios solo pueden ver sus propios privilegios.

iam_role_arn

El rol de IAM al que se le han concedido los privilegios del comando.

cmd_type

El comando para el que se ha concedido acceso. Los valores válidos son los siguientes:

- COPY
- UNLOAD
- EXTERNAL FUNCTION
- CREATE MODEL

### Tipo de retorno

BOOLEAN

Ejemplo

La siguiente consulta confirma que el usuario `reg_user1` tiene el privilegio por el cual el rol `Redshift-S3-Read` puede ejecutar el comando COPY.

```
select has_assumerole_privilege('reg_user1', 'arn:aws:iam::123456789012:role/Redshift-
S3-Read', 'copy');
```

```
has_assumerole_privilege
------------------------
true
(1 row)
```

## HAS_DATABASE_PRIVILEGE

Devuelve `true` si el usuario tiene un privilegio especificado para la base de datos especificada. Para obtener más información acerca de los privilegios, consulte GRANT.

Sintaxis

> **Note**
>
> Esta es una función del nodo principal. Esta función devuelve un error si hace referencia a una tabla creada por usuarios, a una tabla de sistema STL o STV, o a una vista de sistema SVV o SVL.

```
has_database_privilege( [ user, ] database, privilege)
```

Argumentos

usuario

El nombre del usuario para verificar los privilegios de la base de datos. El valor predeterminado es comprobar el usuario actual.

base de datos

La base de datos asociada al privilegio.

privilege

El privilegio que se verificará. Los valores válidos son los siguientes:

- CREATE
- TEMPORARY
- TEMP

Tipo de retorno

Devuelve una cadena CHAR o VARCHAR.

Ejemplo

La siguiente consulta confirma que el usuario GUEST cuenta con el privilegio TEMP en la base de datos TICKIT.

```
select has_database_privilege('guest', 'tickit', 'temp');

has_database_privilege
------------------------
true
(1 row)
```

## HAS_SCHEMA_PRIVILEGE

Devuelve `true` si el usuario tiene un privilegio especificado para el esquema especificado. Para obtener más información acerca de los privilegios, consulte GRANT.

Sintaxis

> **Note**
>
> Esta es una función del nodo principal. Esta función devuelve un error si hace referencia a una tabla creada por usuarios, a una tabla de sistema STL o STV, o a una vista de sistema SVV o SVL.

```
has_schema_privilege( [ user, ] schema, privilege)
```

Argumentos

usuario

El nombre del usuario para verificar los privilegios del esquema. El valor predeterminado es comprobar el usuario actual.

esquema

El esquema asociado al privilegio.

privilege

El privilegio que se verificará. Los valores válidos son los siguientes:

- CREATE
- USAGE
- ALTER
- DROP

Tipo de retorno

Devuelve una cadena CHAR o VARCHAR.

Ejemplo

La siguiente consulta confirma que el usuario GUEST tiene privilegio CREATE en el esquema PUBLIC:

```
select has_schema_privilege('guest', 'public', 'create');

has_schema_privilege
----------------------
true
(1 row)
```

## HAS_TABLE_PRIVILEGE

Devuelve `true` si el usuario tiene un privilegio especificado para la tabla especificada; de lo contrario, devuelve `false`.

Sintaxis

> **Note**
>
> Esta es una función del nodo principal. Esta función devuelve un error si hace referencia a una tabla creada por usuarios, a una tabla de sistema STL o STV, o a una vista de sistema SVV o SVL. Para obtener más información acerca de los privilegios, consulte GRANT.

```
has_table_privilege( [ user, ] table, privilege)
```

Argumentos

usuario

El nombre del usuario para verificar los privilegios de la tabla. El valor predeterminado es comprobar el usuario actual.

tabla

Tabla asociada con el privilegio.

privilege

Privilegio por comprobar. Los valores válidos son los siguientes:

- SELECT
- INSERT
- UPDATE
- DELETE
- DROP
- REFERENCES

Tipo de retorno

BOOLEAN

Ejemplos

La siguiente consulta descubre que el usuario GUEST no cuenta con el privilegio SELECT en la tabla LISTING.

```
select has_table_privilege('guest', 'listing', 'select');

has_table_privilege
--------------------
false
```

La siguiente consulta enumera los privilegios de tabla, incluidos select, insert, update y delete, mediante la salida de las tablas de catálogo pg_tables y pg_user. Se trata solo de un ejemplo. Es posible que deba especificar un nombre de esquema y los nombres de las tablas de su base de datos. Para obtener más información, consulte Consulta de las tablas de catálogos.

```
SELECT
     tablename
     ,usename
     ,HAS_TABLE_PRIVILEGE(users.usename, tablename, 'select') AS sel
     ,HAS_TABLE_PRIVILEGE(users.usename, tablename, 'insert') AS ins
     ,HAS_TABLE_PRIVILEGE(users.usename, tablename, 'update') AS upd
     ,HAS_TABLE_PRIVILEGE(users.usename, tablename, 'delete') AS del
FROM
(SELECT * from pg_tables
WHERE schemaname = 'public' and tablename in ('event','listing')) as tables
,(SELECT * FROM pg_user) AS users;

tablename | usename   |  sel   |  ins  |  upd  | del
----------+-----------+--------+-------+-------+-------
event     |  john     |  true  | true  | true  | true
event     |  sally    |  false | false | false | false
event     |  elsa     |  false | false | false | false
listing   |  john     |  true  | true  | true  | true
listing   |  sally    |  false | false | false | false
listing   |  elsa     |  false | false | false | false
```

La consulta anterior también contiene una combinación cruzada. Para obtener más información, consulte Ejemplos de JOIN. Para consultar tablas que no están en el esquema `public`, elimine la condición `schemaname` de la cláusula WHERE y utilice el siguiente ejemplo antes de la consulta.

```
SET SEARCH_PATH to 'schema_name';
```

# LAST_USER_QUERY_ID

Devuelve el ID de consulta de la última consulta completada del usuario en la sesión actual. Si no se ejecutaron consultas en la sesión actual, last_user_query_id devuelve -1. La función no devuelve el ID de consulta para consultas que se ejecutaron exclusivamente en el nodo principal. Para obtener más información, consulte Funciones específicas del nodo principal.

Sintaxis

```
last_user_query_id()
```

Tipo de retorno

Devuelve un número entero.

Ejemplo

La siguiente consulta devuelve el ID de la última consulta ejecutada por un usuario completada en la sesión actual.

```
select last_user_query_id();
```

Los resultados son los siguientes.

```
last_user_query_id
-----------------------
    5437
(1 row)
```

La siguiente consulta devuelve el ID de consulta y el texto de la última consulta completada ejecutada por un usuario en la sesión actual.

```
select query_id, query_text from sys_query_history where query_id =
 last_user_query_id();
```

Los resultados son los siguientes.

```
 query_id, query_text
```

```
---------
+-----------------------------------------------------------------------------------------------
 5556975 | select last_user_query_id() limit 100 --RequestID=<unique request ID>;
 TraceID=<unique trace ID>
```

## PG_BACKEND_PID

Devuelve el ID de proceso (PID) del proceso del servidor que administra la sesión actual.

> **Note**
>
> El PID no es único en todo el mundo. Se puede volver a usar con el tiempo.

### Sintaxis

```
pg_backend_pid()
```

### Tipo de retorno

Devuelve un número entero.

### Ejemplo

Puede correlacionar PG_BACKEND_PID con las tablas de registro para recuperar información de la sesión actual. Por ejemplo, la siguiente consulta devuelve el ID de consulta y una parte del texto de la consulta para las consultas completadas en la sesión actual.

```
select query, substring(text,1,40)
from stl_querytext
where pid =  PG_BACKEND_PID()
order by query desc;

 query |               substring
-------+----------------------------------------
 14831 | select query, substring(text,1,40) from
 14827 | select query, substring(path,0,80) as pa
 14826 | copy category from 's3://dw-tickit/manif
 14825 | Count rows in target table
 14824 | unload ('select * from category') to 's3
```

```
(5 rows)
```

Puede correlacionar PG_BACKEND_PID con la columna PID en las siguientes tablas de registro (las excepciones se indican entre paréntesis):

- STL_CONNECTION_LOG
- STL_DDLTEXT
- STL_ERROR
- STL_QUERY
- STL_QUERYTEXT
- STL_SESSIONS (procesar)
- STL_TR_CONFLICT
- STL_UTILITYTEXT
- STV_ACTIVE_CURSORS
- STV_INFLIGHT
- STV_LOCKS (lock_owner_pid)
- STV_RECENTS (process_id)

## PG_GET_COLS

Devuelve los metadatos de columna de una definición de vista o tabla.

Sintaxis

```
pg_get_cols('name')
```

Argumentos

name

El nombre de una tabla o una vista de Amazon Redshift. Para obtener más información, consulte Nombres e identificadores.

Tipo de retorno

VARCHAR

### Notas de uso

La función PG_GET_COLS devuelve una fila por cada columna de la definición de vista o tabla. La fila contiene una lista separada por comas con el nombre de esquema, el nombre de relación, el nombre de columna, el tipo de datos y el número de columna. El formato del resultado del SQL depende del cliente SQL utilizado.

### Ejemplos

Los ejemplos siguientes devuelven los resultados de una vista denominada `SALES_VW` en el esquema `public` y una tabla de tablas denominada `sales` en el esquema `mytickit1` creadas por el usuario en la base de datos conectada `dev`.

El siguiente ejemplo devuelve los metadatos de columna de una vista llamada `SALES_VW`.

```
select pg_get_cols('sales_vw');

pg_get_cols
---------------------------------------------------------------
(public,sales_vw,salesid,integer,1)
(public,sales_vw,listid,integer,2)
(public,sales_vw,sellerid,integer,3)
(public,sales_vw,buyerid,integer,4)
(public,sales_vw,eventid,integer,5)
(public,sales_vw,dateid,smallint,6)
(public,sales_vw,qtysold,smallint,7)
(public,sales_vw,pricepaid,"numeric(8,2)",8)
(public,sales_vw,commission,"numeric(8,2)",9)
(public,sales_vw,saletime,"timestamp without time zone",10)
```

El siguiente ejemplo devuelve los metadatos de columna de la vista `SALES_VW` en formato de tabla.

```
select * from pg_get_cols('sales_vw')
cols(view_schema name, view_name name, col_name name, col_type varchar, col_num int);

view_schema | view_name | col_name   | col_type                    | col_num
------------+-----------+------------+-----------------------------+--------
public      | sales_vw  | salesid    | integer                     |       1
public      | sales_vw  | listid     | integer                     |       2
public      | sales_vw  | sellerid   | integer                     |       3
public      | sales_vw  | buyerid    | integer                     |       4
public      | sales_vw  | eventid    | integer                     |       5
public      | sales_vw  | dateid     | smallint                    |       6
```

```
public      | sales_vw  | qtysold    | smallint                    |       7
public      | sales_vw  | pricepaid  | numeric(8,2)                |       8
public      | sales_vw  | commission | numeric(8,2)                |       9
public      | sales_vw  | saletime   | timestamp without time zone |      10
```

El siguiente ejemplo devuelve los metadatos de columna de la tabla SALES en el esquema mytickit1 en formato de tabla.

```
select * from pg_get_cols('"mytickit1"."sales"')
cols(view_schema name, view_name name, col_name name, col_type varchar, col_num int);

view_schema | view_name | col_name   | col_type                    | col_num
------------+-----------+------------+-----------------------------+--------
mytickit1   | sales     | salesid    | integer                     |       1
mytickit1   | sales     | listid     | integer                     |       2
mytickit1   | sales     | sellerid   | integer                     |       3
mytickit1   | sales     | buyerid    | integer                     |       4
mytickit1   | sales     | eventid    | integer                     |       5
mytickit1   | sales     | dateid     | smallint                    |       6
mytickit1   | sales     | qtysold    | smallint                    |       7
mytickit1   | sales     | pricepaid  | numeric(8,2)                |       8
mytickit1   | sales     | commission | numeric(8,2)                |       9
mytickit1   | sales     | saletime   | timestamp without time zone |      10
```

## PG_GET_GRANTEE_BY_IAM_ROLE

Devuelve todos los usuarios y los grupos a los que se ha otorgado un rol de IAM especificado.

Sintaxis

```
pg_get_grantee_by_iam_role('iam_role_arn')
```

Argumentos

iam_role_arn

El rol de IAM para el que se devuelven los usuarios y los grupos a los que se ha otorgado este rol.

Tipo de retorno

VARCHAR

### Notas de uso

La función PG_GET_GRANTEE_BY_IAM_ROLE devuelve una fila por cada usuario o grupo. Cada fila contiene el nombre del beneficiario, el tipo de beneficiario y el privilegio otorgado. Los valores posibles para el tipo de beneficiario son p para público, u para usuario y g para grupo.

Debe ser un superusuario para utilizar esta función.

### Ejemplo

En el siguiente ejemplo, se indica que el rol de IAM `Redshift-S3-Write` se otorga a `group1` y `reg_user1`. Los usuarios de `group_1` pueden especificar el rol solo para las operaciones COPY, y el usuario `reg_user1` puede especificar el rol solo para realizar operaciones UNLOAD.

```
select pg_get_grantee_by_iam_role('arn:aws:iam::123456789012:role/Redshift-S3-Write');
```

```
  pg_get_grantee_by_iam_role
-----------------------------
 (group_1,g,COPY)
 (reg_user1,u,UNLOAD)
```

En el siguiente ejemplo de la función PG_GET_GRANTEE_BY_IAM_ROLE, el resultado recibe el formato de tabla.

```
select grantee, grantee_type, cmd_type FROM
 pg_get_grantee_by_iam_role('arn:aws:iam::123456789012:role/Redshift-S3-Write')
 res_grantee(grantee text, grantee_type text, cmd_type text) ORDER BY 1,2,3;
```

```
  grantee  | grantee_type | cmd_type
-----------+--------------+----------
 group_1   | g            | COPY
 reg_user1 | u            | UNLOAD
```

## PG_GET_IAM_ROLE_BY_USER

Devuelve todos los roles de IAM y los privilegios de comando otorgados a un usuario.

### Sintaxis

```
pg_get_iam_role_by_user('name')
```

Argumentos

name

El nombre del usuario al que se devuelven los roles de IAM.

Tipo de retorno

VARCHAR

Notas de uso

La función PG_GET_IAM_ROLE_BY_USER devuelve una fila por cada conjunto de roles y privilegios de comando. La fila contiene una lista separada por comas con el nombre de usuario, el rol de IAM y el comando.

Un valor `default` en el resultado indica que el usuario puede especificar cualquier rol disponible para ejecutar el comando mostrado.

Debe ser un superusuario para utilizar esta función.

Ejemplo

En el siguiente ejemplo, se indica que el usuario `reg_user1` puede especificar cualquier rol de IAM disponible para realizar operaciones COPY. El usuario también puede especificar el rol `Redshift-S3-Write` para las operaciones UNLOAD.

```
select pg_get_iam_role_by_user('reg_user1');
```

```
                            pg_get_iam_role_by_user
-------------------------------------------------------------------------------
 (reg_user1,default,COPY)
 (reg_user1,arn:aws:iam::123456789012:role/Redshift-S3-Write,COPY|UNLOAD)
```

En el siguiente ejemplo de la función PG_GET_IAM_ROLE_BY_USER, el resultado recibe el formato de tabla.

```
select username, iam_role, cmd FROM pg_get_iam_role_by_user('reg_user1')
 res_iam_role(username text, iam_role text, cmd text);
```

```
 username  |                    iam_role                     | cmd
-----------+-------------------------------------------------+------
 reg_user1 | default                                         | None
 reg_user1 | arn:aws:iam::123456789012:role/Redshift-S3-Read | COPY
```

## PG_GET_LATE_BINDING_VIEW_COLS

Devuelve los metadatos de columna de todas las vistas de enlace de tiempo de ejecución de la base de datos. Para obtener más información, consulte Vistas de enlace en tiempo de ejecución

Sintaxis

```
pg_get_late_binding_view_cols()
```

Tipo de retorno

VARCHAR

Notas de uso

La función PG_GET_LATE_BINDING_VIEW_COLS devuelve una fila por cada columna de las vistas de enlace de tiempo de ejecución. La fila contiene una lista separada por comas con el nombre de esquema, el nombre de relación, el nombre de columna, el tipo de datos y el número de columna.

Ejemplo

El siguiente ejemplo devuelve los metadatos de columna de todas las vistas de enlace de tiempo de ejecución.

```
select pg_get_late_binding_view_cols();

pg_get_late_binding_view_cols
------------------------------------------------------------
(public,myevent,eventname,"character varying(200)",1)
(public,sales_lbv,salesid,integer,1)
(public,sales_lbv,listid,integer,2)
(public,sales_lbv,sellerid,integer,3)
(public,sales_lbv,buyerid,integer,4)
(public,sales_lbv,eventid,integer,5)
(public,sales_lbv,dateid,smallint,6)
```

```
(public,sales_lbv,qtysold,smallint,7)
(public,sales_lbv,pricepaid,"numeric(8,2)",8)
(public,sales_lbv,commission,"numeric(8,2)",9)
(public,sales_lbv,saletime,"timestamp without time zone",10)
(public,event_lbv,eventid,integer,1)
(public,event_lbv,venueid,smallint,2)
(public,event_lbv,catid,smallint,3)
(public,event_lbv,dateid,smallint,4)
(public,event_lbv,eventname,"character varying(200)",5)
(public,event_lbv,starttime,"timestamp without time zone",6)
```

El siguiente ejemplo devuelve los metadatos de columna de todas las vistas de enlace de tiempo de ejecución en formato de tabla.

```
select * from pg_get_late_binding_view_cols() cols(view_schema name, view_name name,
 col_name name, col_type varchar, col_num int);
view_schema | view_name | col_name   | col_type                    | col_num
------------+-----------+------------+-----------------------------+--------
public      | sales_lbv | salesid    | integer                     |       1
public      | sales_lbv | listid     | integer                     |       2
public      | sales_lbv | sellerid   | integer                     |       3
public      | sales_lbv | buyerid    | integer                     |       4
public      | sales_lbv | eventid    | integer                     |       5
public      | sales_lbv | dateid     | smallint                    |       6
public      | sales_lbv | qtysold    | smallint                    |       7
public      | sales_lbv | pricepaid  | numeric(8,2)                |       8
public      | sales_lbv | commission | numeric(8,2)                |       9
public      | sales_lbv | saletime   | timestamp without time zone |      10
public      | event_lbv | eventid    | integer                     |       1
public      | event_lbv | venueid    | smallint                    |       2
public      | event_lbv | catid      | smallint                    |       3
public      | event_lbv | dateid     | smallint                    |       4
public      | event_lbv | eventname  | character varying(200)      |       5
public      | event_lbv | starttime  | timestamp without time zone |       6
```

## PG_GET_SESSION_ROLES

Devuelve los roles de sesión del usuario que ha iniciado sesión actualmente. Los roles de sesión de un usuario son los grupos definidos por un proveedor de identidades (IdP) para el usuario que ha iniciado sesión. Por ejemplo, un proveedor de identidades (IdP) como Microsoft Azure Active Directory (Azure AD) verifica la identidad del usuario y proporciona los grupos externos de los que forma parte el usuario durante el proceso de inicio de sesión del usuario. Estos grupos externos se

transforman en roles de Amazon Redshift y están disponibles durante la sesión actual. Estos roles se denominan roles de sesión. Un administrador puede conceder privilegios a un rol de sesión similar a otros roles de Amazon Redshift. Para obtener información acerca del uso de roles, consulte Control de acceso basado en roles (RBAC). Para obtener información sobre la administración de identidades con un proveedor de identidades (IdP), consulte Federación de proveedor de identidades (IdP) nativo para Amazon Redshift en la Guía de administración de Amazon Redshift.

Para ver los roles definidos en el catálogo de Amazon Redshift, consulte la vista del sistema SVV_ROLES.

Sintaxis

```
pg_get_session_roles()
```

Tipo de retorno

Un conjunto de filas que consta de dos valores. El primer valor tiene dos partes separadas por dos puntos (:) que contiene un `idp-namespace:role-name`. `idp-namespace` es el espacio de nombres del proveedor de identidades (IdP). `role-name` es el nombre del grupo externo del proveedor de identidades (IdP). El segundo valor contiene un `role-id` que es el identificador del rol.

Notas de uso

La función `PG_GET_SESSION_ROLES` devuelve una fila por cada rol de sesión devuelto.

Ejemplos

El siguiente ejemplo devuelve una fila para cada rol del IdP de Azure Active Directory. Las columnas devueltas se convierten en `sess_roles` con columnas `name` y `roleid`. Cada `name` consta del espacio de nombres de Azure Active Directory y un nombre de grupo en Azure Active Directory.

```
SELECT * FROM pg_get_session_roles() AS sess_roles(name name, roleid integer);

name                   roleid
--------------------------------
my_aad:test_group_1   106204
my_aad:test_group_2   106205
my_aad:test_group_3   106206
my_aad:test_group_4   106207
my_aad:test_group_5   106208
```

El siguiente ejemplo devuelve una fila para cada grupo de IAM del que es miembro el usuario de IAM que ha iniciado sesión actualmente. Las columnas devueltas se convierten en `sess_roles` con columnas `name` y `roleid`. Cada `name` consta del espacio de nombres de IAM y el nombre del grupo de IAM.

```
SELECT * FROM pg_get_session_roles() AS sess_roles(name name, roleid integer);

name                  roleid
-------------------------------
IAM:myGroup           110332
```

## PG_LAST_COPY_COUNT

Devuelve la cantidad de filas cargadas por el último comando COPY ejecutado en la sesión actual. PG_LAST_COPY_COUNT se actualiza con el último COPY ID, que es el ID de consulta del último COPY que comenzó el proceso de carga, aun cuando la carga haya fallado. El ID de consulta y el COPY ID se actualizan cuando el comando COPY comienza el proceso de carga.

Si COPY falla por un error de sintaxis o porque no hay privilegios suficientes, el COPY ID no se actualiza y PG_LAST_COPY_COUNT devuelve el conteo del comando COPY anterior. Si no se ejecutaron comandos COPY en la sesión actual, o si el último COPY dio error durante la carga, PG_LAST_COPY_COUNT devuelve 0. Para obtener más información, consulte PG_LAST_COPY_ID.

Sintaxis

```
pg_last_copy_count()
```

Tipo de retorno

Devuelve BIGINT.

Ejemplo

La siguiente consulta devuelve la cantidad de filas cargadas por el último comando COPY ejecutado en la sesión actual.

```
select pg_last_copy_count();

pg_last_copy_count
--------------------
             192497
```

```
(1 row)
```

## PG_LAST_COPY_ID

Devuelve el ID de consulta del último comando COPY que se haya completado en la sesión actual. Si no se ejecutaron comandos COPY en la sesión actual, PG_LAST_COPY_ID devuelve -1.

El valor para PG_LAST_COPY_ID se actualiza cuando el comando COPY comienza el proceso de carga. Si el COPY falla por datos de carga no válidos, el COPY ID se actualiza, por lo que puede usar PG_LAST_COPY_ID cuando consulte la tabla STL_LOAD_ERRORS. Si la transacción COPY se revierte, el COPY ID no se actualiza.

El COPY ID no se actualiza si el comando COPY falla por un error que ocurre antes de que comience el proceso de carga, como un error de sintaxis, error de acceso, credenciales no válidas o privilegios insuficientes. El COPY ID no se actualiza si el COPY falla durante el paso de análisis de compresión, que comienza después de una conexión con éxito, pero antes de la carga de datos.

Sintaxis

```
pg_last_copy_id()
```

Tipo de retorno

Devuelve un número entero.

Ejemplo

La siguiente consulta devuelve el ID de consulta para el último comando COPY ejecutado en la sesión actual.

```
select pg_last_copy_id();

pg_last_copy_id
---------------
          5437
(1 row)
```

La siguiente consulta combina STL_LOAD_ERRORS y STL_LOADERROR_DETAIL para ver los errores de detalles que ocurrieron durante la carga más reciente en la sesión actual:

```
select d.query, substring(d.filename,14,20),
```

```
d.line_number as line,
substring(d.value,1,16) as value,
substring(le.err_reason,1,48) as err_reason
from stl_loaderror_detail d, stl_load_errors le
where d.query = le.query
and d.query = pg_last_copy_id();

 query |    substring      | line |  value   |                    err_reason
-------+-------------------+------+----------
+-----------------------------------------------
    558| allusers_pipe.txt |  251 | 251      | String contains invalid or unsupported
 UTF8 code
    558| allusers_pipe.txt |  251 | ZRU29FGR | String contains invalid or unsupported
 UTF8 code
    558| allusers_pipe.txt |  251 | Kaitlin  | String contains invalid or unsupported
 UTF8 code
    558| allusers_pipe.txt |  251 | Walter   | String contains invalid or unsupported
 UTF8 code
```

## PG_LAST_UNLOAD_ID

Devuelve el ID de consulta del último comando UNLOAD que se haya completado en la sesión actual. Si no se ejecutaron comandos UNLOAD en la sesión actual, PG_LAST_UNLOAD_ID devuelve -1.

El valor de PG_LAST_UNLOAD_ID se actualiza cuando el comando UNLOAD comienza el proceso de carga. Si el comando UNLOAD da un error por datos de carga no válidos, se actualiza el UNLOAD ID para que pueda usarlo para sus investigaciones. Si la transacción UNLOAD se revierte, el UNLOAD ID no se actualiza.

El UNLOAD ID no se actualiza si el comando UNLOAD produce un error que ocurre antes de que comience el proceso de carga, como un error de sintaxis, error de acceso, credenciales no válidas o privilegios insuficientes.

Sintaxis

```
PG_LAST_UNLOAD_ID()
```

Tipo de retorno

Devuelve un número entero.

Ejemplo

La siguiente consulta devuelve el ID de consulta para el último comando UNLOAD ejecutado en la sesión actual.

```
select PG_LAST_UNLOAD_ID();

PG_LAST_UNLOAD_ID
---------------
          5437
(1 row)
```

## PG_LAST_QUERY_ID

Devuelve el ID de consulta de la última consulta que se haya completado en la sesión actual. Si no se ejecutaron consultas en la sesión actual, PG_LAST_QUERY_ID devuelve -1. PG_LAST_QUERY_ID no devuelve el ID de consulta para consultas que se ejecutaron exclusivamente en el nodo principal. Para obtener más información, consulte Funciones específicas del nodo principal.

Sintaxis

```
pg_last_query_id()
```

Tipo de retorno

Devuelve un número entero.

Ejemplo

La siguiente consulta devuelve el ID de la última consulta que se haya completado en la sesión actual.

```
select pg_last_query_id();
```

Los resultados son los siguientes.

```
pg_last_query_id
----------------
           5437
```

```
(1 row)
```

La siguiente consulta devuelve el ID de consulta y el texto de la última consulta que se haya realizado en la sesión actual.

```
select query, trim(querytxt) as sqlquery
from stl_query
where query = pg_last_query_id();
```

Los resultados son los siguientes.

```
query | sqlquery
------+-------------------------------------------------
 5437 | select name, loadtime from stl_file_scan where loadtime > 1000000;
(1 rows)
```

## PG_LAST_UNLOAD_COUNT

Devuelve la cantidad de filas descargadas por el último comando UNLOAD que se haya completado en la sesión actual. PG_LAST_UNLOAD_COUNT se actualiza con el último ID de consulta del último UNLOAD, aun cuando la operación haya fallado. El ID de consulta se actualiza cuando se completa UNLOAD. Si UNLOAD falla por un error de sintaxis o porque no hay privilegios suficientes, PG_LAST_UNLOAD_COUNT devuelve el conteo del comando UNLOAD anterior. Si no se completaron comandos UNLOAD en la sesión actual o si el último UNLOAD dio error durante la operación de descarga, PG_LAST_UNLOAD_COUNT devuelve 0.

Sintaxis

```
pg_last_unload_count()
```

Tipo de retorno

Devuelve BIGINT.

Ejemplo

La siguiente consulta devuelve la cantidad de filas descargadas por el último comando UNLOAD ejecutado en la sesión actual.

```
select pg_last_unload_count();
```

```
pg_last_unload_count
--------------------
              192497
(1 row)
```

## Función SLICE_NUM

Devuelve un valor entero correspondiente al número de sección en el clúster donde se ubican los datos de una fila. SLICE_NUM no toma parámetros.

### Sintaxis

```
SLICE_NUM()
```

### Tipo de retorno

La función SLICE_NUM devuelve un valor entero.

### Ejemplos

En el siguiente ejemplo, se muestra qué secciones tienen datos para las primeras diez filas de EVENT en la tabla EVENTS:

```
select distinct eventid, slice_num() from event order by eventid limit 10;

 eventid | slice_num
---------+-----------
       1 |         1
       2 |         2
       3 |         3
       4 |         0
       5 |         1
       6 |         2
       7 |         3
       8 |         0
       9 |         1
      10 |         2
(10 rows)
```

El siguiente ejemplo devuelve un código (10000) para mostrar que una consulta sin instrucción FROM se ejecuta en el nodo principal:

```
select slice_num();
slice_num
-----------
10000
(1 row)
```

## USER

Sinónimo de CURRENT_USER. Consulte CURRENT_USER.

## ROLE_IS_MEMBER_OF

Devuelve el valor true (verdadero) si el rol es miembro de otro rol. Los superusuarios pueden comprobar la pertenencia de todos los roles. Los usuarios normales que tengan el permiso ACCESS SYSTEM TABLE pueden comprobar la pertenencia de todos los usuarios. En caso contrario, los usuarios normales solo pueden comprobar los roles a los que tengan acceso. Amazon Redshift falla si los roles proporcionados no existen o el usuario actual no tiene acceso al rol.

Sintaxis

```
role_is_member_of( role_name,  granted_role_name)
```

Argumentos

role_name

Nombre del rol

granted_role_name

Nombre del rol concedido.

Tipo de retorno

Devuelve un BOOLEANO.

Ejemplo

La siguiente consulta confirma que el rol no es miembro de role1 ni de role2.

```
SELECT role_is_member_of('role1', 'role2');
```

```
 role_is_member_of
-------------------
             False
```

## USER_IS_MEMBER_OF

Devuelve true (verdadero) si el usuario es miembro de un rol o grupo. Los superusuarios pueden comprobar la pertenencia de todos los usuarios. Los usuarios normales que sean miembros del rol sys:secadmin o sys:superuser pueden comprobar la pertenencia de todos los usuarios. En caso contrario, los usuarios normales solo pueden comprobarse a sí mismos. Amazon Redshift envía un error si las identidades proporcionadas no existen o el usuario actual no tiene acceso al rol.

Sintaxis

```
user_is_member_of( user_name,  role_name | group_name)
```

Argumentos

user_name

El nombre del usuario.

role_name

Nombre del rol

group_name

El nombre del grupo.

Tipo de retorno

Devuelve un BOOLEANO.

Ejemplo

La siguiente consulta confirma que el usuario no es miembro de role1.

```
SELECT user_is_member_of('reguser', 'role1');

 user_is_member_of
-------------------
```

```
            False
```

## VERSION

La función VERSION devuelve detalles acerca de la versión instalada en la actualidad, con información de la versión específica de Amazon Redshift al final.

> **Note**
>
> Esta es una función del nodo principal. Esta función devuelve un error si hace referencia a una tabla creada por usuarios, a una tabla de sistema STL o STV, o a una vista de sistema SVV o SVL.

### Sintaxis

```
VERSION()
```

### Tipo de retorno

Devuelve una cadena CHAR o VARCHAR.

### Ejemplos

En el siguiente ejemplo, se muestra la información de versión del clúster actual:

```
select version();
```

```
 version
 ----------------------------------------------------------------------------------------------------------------

 PostgreSQL 8.0.2 on i686-pc-linux-gnu, compiled by GCC gcc (GCC) 3.4.2 20041017 (Red
 Hat 3.4.2-6.fc3), Redshift 1.0.12103
```

Dónde `1.0.12103` es el número de versión del clúster.

> **Note**
>
> Para hacer que el clúster se actualice a la última versión, ajuste el periodo de mantenimiento.

# Palabras reservadas

A continuación, se ofrece una lista de las palabras reservadas de Amazon Redshift. Puede utilizar las palabras reservadas con identificadores delimitados (comillas dobles).

> **Note**
>
> Aunque START y CONNECT no son palabras reservadas, utilice identificadores delimitados o AS si utiliza START y CONNECT como alias de tabla en su consulta para evitar errores en tiempo de ejecución.

Para obtener más información, consulte Nombres e identificadores.

```
AES128
AES256
ALL
ALLOWOVERWRITE
ANALYSE
ANALYZE
AND
ANY
ARRAY
AS
ASC
AUTHORIZATION
AZ64
BACKUP
BETWEEN
BINARY
BLANKSASNULL
BOTH
BYTEDICT
BZIP2
CASE
CAST
CHECK
COLLATE
COLUMN
CONSTRAINT
CREATE
CREDENTIALS
```

```
CROSS
CURRENT_DATE
CURRENT_TIME
CURRENT_TIMESTAMP
CURRENT_USER
CURRENT_USER_ID
DEFAULT
DEFERRABLE
DEFLATE
DEFRAG
DELTA
DELTA32K
DESC
DISABLE
DISTINCT
DO
ELSE
EMPTYASNULL
ENABLE
ENCODE
ENCRYPT
ENCRYPTION
END
EXCEPT
EXPLICIT
FALSE
FOR
FOREIGN
FREEZE
FROM
FULL
GLOBALDICT256
GLOBALDICT64K
GRANT
GROUP
GZIP
HAVING
IDENTITY
IGNORE
ILIKE
IN
INITIALLY
INNER
INTERSECT
```

```
INTERVAL
INTO
IS
ISNULL
JOIN
LEADING
LEFT
LIKE
LIMIT
LOCALTIME
LOCALTIMESTAMP
LUN
LUNS
LZO
LZOP
MINUS
MOSTLY16
MOSTLY32
MOSTLY8
NATURAL
NEW
NOT
NOTNULL
NULL
NULLS
OFF
OFFLINE
OFFSET
OID
OLD
ON
ONLY
OPEN
OR
ORDER
OUTER
OVERLAPS
PARALLEL
PARTITION
PERCENT
PERMISSIONS
PIVOT
PLACING
PRIMARY
```

```
RAW
READRATIO
RECOVER
REFERENCES
REJECTLOG
RESORT
RESPECT
RESTORE
RIGHT
SELECT
SESSION_USER
SIMILAR
SNAPSHOT
SOME
SYSDATE
SYSTEM
TABLE
TAG
TDES
TEXT255
TEXT32K
THEN
TIMESTAMP
TO
TOP
TRAILING
TRUE
TRUNCATECOLUMNS
UNION
UNIQUE
UNNEST
UNPIVOT
USER
USING
VERBOSE
WALLET
WHEN
WHERE
WITH
WITHOUT
```

# Referencia de las tablas y vistas de sistema

Amazon Redshift dispone de muchas tablas y vistas de sistema que contienen información acerca de cómo funciona el sistema. Puede consultar estas tablas y vistas de sistema de la misma forma que lo haría con cualquier otra tabla de bases de datos. En esta sección se muestran algunas consultas y explicaciones de ejemplo sobre las tablas de sistema:

- Cómo se generan las diferentes vistas y tablas de sistema.
- Qué tipos de información se puede obtener de estas tablas.
- Cómo unir tablas de sistema de Amazon Redshift a tablas de catálogos
- Cómo administrar el crecimiento de los archivos de registro de las tablas de sistema.

Algunas tablas de sistema solo pueden ser utilizadas por el personal de AWS con fines de diagnóstico. En las siguientes secciones se explican las tablas de sistema que los administradores del sistema u otros usuarios de la base de datos pueden consultar para obtener información útil.

> **Note**
>
> Las tablas de sistema no están incluidas en los backups de clústeres manuales o automáticos (instantáneas). Las vistas del sistema STL retienen siete días del historial de registros. La retención de registros no requiere ninguna acción por parte del cliente, pero si desea almacenar datos de registro durante más de siete días, deberá copiarlos periódicamente en otras tablas o descargarlos en Amazon S3.

Temas


- Tipos de tablas y vistas de sistema
- Visibilidad de datos en las tablas y vistas de sistema
- Migración de consultas aprovisionadas solo a consultas de vista de supervisión de SYS
- Mejora del seguimiento de los identificadores de consultas mediante las vistas de monitoreo SYS
- Identificadores de consulta, proceso y sesión de tablas del sistema
- Vistas de metadatos SVV
- Vistas de monitoreo de SYS
- Asignación de la vista de sistema para migrar a las vistas de supervisión SYS

- Monitoreo del sistema (solo aprovisionado)
- Tablas de catálogos de sistema

# Tipos de tablas y vistas de sistema

Existen varios tipos de tablas y vistas del sistema:

- Las vistas SVV contienen información sobre los objetos de la base de datos con referencias a tablas STV transitorias.
- Las vistas SYS se utilizan para monitorizar el uso de consultas y cargas de trabajo de los clústeres y grupos de trabajo sin servidor aprovisionados.
- Las vistas STL se generan a partir de los registros que se han almacenado en el disco para proporcionar un historial del sistema.
- Las tablas de STV son tablas de sistema virtuales que tienen instantáneas de los datos actuales del sistema. Están basadas en los datos transitorios en memoria y no se almacenan en los registros en disco o en las tablas normales.
- Las vistas SVCS proporcionan información sobre las consultas de los clústeres principal y de escalado de simultaneidad.
- Las vistas SVL proporcionan detalles sobre las consultas en los clústeres principales.

Las tablas y vistas de sistema no utilizan el mismo modelo de consistencia que las tablas normales. Es importante tener en cuenta este problema cuando se las consulta, en especial con las tablas STV y las vistas SVV. Por ejemplo, en el caso de una tabla t1 normal con una columna c1, se espera que la siguiente consulta no devuelva filas:

```
select * from t1
where c1 > (select max(c1) from t1)
```

No obstante, la siguiente consulta con una tabla de sistema puede devolver filas:

```
select * from stv_exec_state
where currenttime > (select max(currenttime) from stv_exec_state)
```

La razón por la que esta consulta puede devolver filas es que currenttime es transitorio y las dos referencias de la consulta podrían no devolver el mismo valor cuando se evalúan.

Por otra parte, la siguiente consulta puede no devolver filas:

```
select * from stv_exec_state
where currenttime = (select max(currenttime) from stv_exec_state)
```

## Visibilidad de datos en las tablas y vistas de sistema

Existen dos clases de visibilidad para los datos de las tablas y vistas de sistema: visibilidad para los usuarios y visibilidad para los superusuarios.

Los usuarios con privilegios de superusuario son los únicos que pueden ver los datos de las tablas visibles para los superusuarios. Los usuarios normales pueden ver los datos de las tablas visibles para los usuarios. Para brindar a un usuario normal acceso a tablas visibles para los superusuarios, concédale el privilegio SELECT de esa tabla. Para obtener más información, consulte GRANT.

En las tablas visibles para los usuarios, de forma predeterminada, los usuarios normales no pueden ver las filas que generan otros usuarios. Si a un usuario normal se le concede SYSLOG ACCESS UNRESTRICTED, ese usuario podrá ver todas las filas de las tablas visibles para los usuarios, incluidas las que haya generado otro usuario. Para obtener más información, consulte ALTER USER o CREAR USUARIO. Todas las filas de SVV_TRANSACTIONS son visibles para todos los usuarios. Para obtener más información sobre la visibilidad de los datos, consulte el artículo de la base de conocimientos AWS re:Post ¿Cómo puedo conceder permiso a los usuarios normales de bases de datos de Amazon Redshift para ver los datos de las tablas del sistema de otros usuarios de mi clúster?.

En el caso de las vistas de metadatos, Amazon Redshift no permite la visibilidad a los usuarios a los que se ha concedido SYSLOG ACCESS UNRESTRICTED.

> **Note**
>
> Al brindar a un usuario acceso sin restricciones a las tablas del sistema, le proporciona la visibilidad necesaria para ver los datos generados por otros usuarios. Por ejemplo, STL_QUERY y STL_QUERY_TEXT contienen todo el texto de las instrucciones INSERT, UPDATE y DELETE, que podrían contener datos confidenciales generados por los usuarios.

Un super usuario puede ver todas las filas de todas las tablas. Para brindar a un usuario normal acceso a una tabla visible para los superusuarios, concédale el privilegio GRANT SELECT de esta tabla.

### Filtrado de consultas generadas por el sistema

Las tablas y las vistas de sistema relacionadas con las consultas, como SVL_QUERY_SUMMARY, SVL_QLOG y otras, suelen contener una gran cantidad de instrucciones generadas automáticamente que Amazon Redshift utiliza para supervisar el estado de la base de datos. Estas consultas generadas por el sistema son visibles para un super usuario, pero raramente son útiles. Para filtrarlas cuando se selecciona de una tabla de sistema o vista de sistema que utiliza la columna `userid`, agregue la condición `userid > 1` a la cláusula WHERE. Por ejemplo:

```
select * from svl_query_summary where userid > 1
```

## Migración de consultas aprovisionadas solo a consultas de vista de supervisión de SYS

### Migración de clústeres aprovisionados a Amazon Redshift sin servidor

Si va a migrar un clúster aprovisionado a Amazon Redshift sin servidor, es posible que tenga consultas que utilicen las siguientes vistas del sistema, que solo almacenan datos de clústeres aprovisionados.

- Todas las vistas STL
- Todas las vistas STV
- Todas las vistas SVCS
- Todas las vistas SVL
- Algunas vistas SVV
  - Para obtener una lista completa de las vistas SVV no compatibles en Amazon Redshift sin servidor, consulte la lista que aparece al final de Supervisión de consultas y cargas de trabajo con Amazon Redshift sin servidor en la Guía de administración de Amazon Redshift.

Para seguir utilizando sus consultas, reajústelas para utilizar las columnas definidas en las vistas de monitorización SYS que correspondan a las columnas de sus vistas aprovisionadas solo. Para ver la relación de asignación entre las vistas de solo aprovisionamiento y las vistas de supervisión de SYS, vaya a Asignación de la vista de sistema para migrar a las vistas de supervisión SYS

## Actualización de consultas mientras se permanece en un clúster aprovisionado

Si no va a migrar a Amazon Redshift sin servidor, es posible que desee actualizar las consultas existentes. Las vistas de monitorización SYS están diseñadas para facilitar su utilización y reducir su complejidad, proporcionando una completa gama de métricas para monitorear y solucionar problemas de forma eficaz. Mediante vistas SYS como SYS_QUERY_HISTORY y SYS_QUERY_DETAIL que consolidan la información de varias vistas aprovisionadas solo, puede agilizar sus consultas.

# Mejora del seguimiento de los identificadores de consultas mediante las vistas de monitoreo SYS

Las vistas de monitoreo SYS, como SYS_QUERY_HISTORY y SYS_QUERY_DETAIL, contienen la columna query_id, que incluye el identificador de las consultas de los usuarios. Del mismo modo, las vistas solo aprovisionadas, como STL_QUERY y SVL_QLOG, contienen la columna de consulta, que también incluye los identificadores de consultas. Sin embargo, los identificadores de consultas registrados en las vistas del sistema SYS son diferentes de los registrados en las vistas únicamente aprovisionadas.

La diferencia entre los valores de la columna query_id de las vistas SYS y los valores de las columnas de consulta de las vistas únicamente aprovisionadas es la siguiente:

- En las vistas SYS, la columna query_id registra las consultas enviadas por los usuarios en su formato original. El optimizador de Amazon Redshift puede dividirlas en consultas secundarias para mejorar el rendimiento, pero si ejecuta una sola consulta esta seguirá teniendo una sola fila en SYS_QUERY_HISTORY. Si desea ver las consultas secundarias individuales, puede consultarlas en SYS_QUERY_DETAIL.
- En las vistas únicamente aprovisionadas, la columna de consultas registra las consultas en el nivel de consulta secundaria. Si el optimizador de Amazon Redshift reescribe la consulta original en varias consultas secundarias, habrá varias filas en STL_QUERY con valores de identificador de consulta diferentes para una sola consulta que ejecute.

Cuando migre sus consultas de monitorep y diagnóstico de vistas únicamente aprovisionadas a vistas SYS, tenga en cuenta esta diferencia y edite las consultas en consecuencia. Para obtener

más información sobre cómo procesa Amazon Redshift las consultas, consulte Flujo de trabajo de planificación y ejecución de consultas.

## Ejemplo

Para ver un ejemplo de cómo Amazon Redshift registra las consultas de forma diferente en las vistas de supervisión de SYS y de solo aprovisionamiento, consulte la siguiente consulta de ejemplo. Esta es la consulta escrita tal y como la ejecutaría en Amazon Redshift.

```
SELECT
  s_name
  , COUNT(*) AS numwait
FROM
  supplier,
  lineitem l1,
  orders,
  nation
WHERE    s_suppkey = l1.l_suppkey
         AND o_orderkey = l1.l_orderkey
         AND o_orderstatus = 'F'
         AND l1.l_receiptdate > l1.l_commitdate
         AND EXISTS (SELECT
                       *
                     FROM
                       lineitem l2
                     WHERE  l2.l_orderkey = l1.l_orderkey
                            AND l2.l_suppkey <> l1.l_suppkey )
         AND NOT EXISTS (SELECT
                           *
                         FROM
                           lineitem l3
                         WHERE  l3.l_orderkey = l1.l_orderkey
                                AND l3.l_suppkey <> l1.l_suppkey
                                AND l3.l_receiptdate > l3.l_commitdate )
         AND s_nationkey = n_nationkey
         AND n_name = 'UNITED STATES'
GROUP BY
  s_name
ORDER BY
  numwait DESC
  , s_name LIMIT 100;
```

Internamente, el optimizador de consultas de Amazon Redshift reescribe la consulta anterior enviada por el usuario en cinco consultas secundarias.

La primera consulta secundaria crea una tabla temporal para materializar una subconsulta.

```
CREATE TEMP TABLE volt_tt_606590308b512(l_orderkey
                                       , l_suppkey
                                       , s_name   ) AS SELECT
                                                       l1.l_orderkey
                                                       , l1.l_suppkey
                                                       , public.supplier.s_name
                                                     FROM
                                                       public.lineitem AS l1,
                                                       public.nation,
                                                       public.orders,
                                                       public.supplier
                                                     WHERE  l1.l_commitdate <
 l1.l_receiptdate
                                                            AND l1.l_orderkey =
 public.orders.o_orderkey
                                                            AND l1.l_suppkey =
 public.supplier.s_suppkey
                                                            AND public.nation.n_name
 = 'UNITED STATES'::CHAR(8)
                                                            AND
 public.nation.n_nationkey = public.supplier.s_nationkey
                                                            AND
 public.orders.o_orderstatus = 'F'::CHAR(1);
```

La segunda consulta secundaria recopila estadísticas de la tabla temporal.

```
padb_fetch_sample: select count(*) from volt_tt_606590308b512;
```

La tercera consulta secundaria crea otra tabla temporal para materializar otra subconsulta, que hace referencia a la tabla temporal creada anteriormente.

```
CREATE TEMP TABLE volt_tt_606590308c2ef(l_orderkey
                                       , l_suppkey) AS (SELECT

 volt_tt_606590308b512.l_orderkey
                                                          ,
 volt_tt_606590308b512.l_suppkey
```

```
                                                    FROM
                                                      public.lineitem AS l2,
                                                      volt_tt_606590308b512
                                                    WHERE  l2.l_suppkey <>
 volt_tt_606590308b512.l_suppkey
                                                           AND l2.l_orderkey =
 volt_tt_606590308b512.l_orderkey)
                                                           EXCEPT distinct (SELECT
 volt_tt_606590308b512.l_orderkey, volt_tt_606590308b512.l_suppkey
                                                           FROM public.lineitem AS
 l3, volt_tt_606590308b512
                                                           WHERE l3.l_commitdate <
 l3.l_receiptdate
                                                             AND l3.l_suppkey <>
 volt_tt_606590308b512.l_suppkey
                                                             AND l3.l_orderkey =
 volt_tt_606590308b512.l_orderkey);
```

La segunda consulta secundaria vuelve a recopilar las estadísticas de la tabla temporal.

```
padb_fetch_sample: select count(*) from volt_tt_606590308c2ef
```

La última consulta secundaria utiliza las tablas temporales creadas anteriormente para generar el resultado.

```
SELECT
  volt_tt_606590308b512.s_name AS s_name
  , COUNT(*) AS numwait
FROM
  volt_tt_606590308b512,
  volt_tt_606590308c2ef
WHERE    volt_tt_606590308b512.l_orderkey = volt_tt_606590308c2ef.l_orderkey
         AND volt_tt_606590308b512.l_suppkey = volt_tt_606590308c2ef.l_suppkey
GROUP BY
  1
ORDER BY
  2 DESC
  , 1 ASC LIMIT 100;
```

En la vista de sistema STL_QUERY, únicamente aprovisionada, Amazon Redshift registra cinco filas en el nivel de consulta secundaria, de la siguiente manera:

```
SELECT userid, xid, pid, query, querytxt::varchar(100);
FROM stl_query
WHERE xid = 48237350
ORDER BY xid, starttime;

 userid |   xid    |    pid     |  query   |
    querytxt
--------+----------+------------+----------
+-----------------------------------------------------------------------------------------------------------
    101 | 48237350 | 1073840810 | 12058151 | CREATE TEMP TABLE
 volt_tt_606590308b512(l_orderkey, l_suppkey, s_name) AS SELECT l1.l_orderkey, l1.l
    101 | 48237350 | 1073840810 | 12058152 | padb_fetch_sample: select count(*) from
 volt_tt_606590308b512
    101 | 48237350 | 1073840810 | 12058156 | CREATE TEMP TABLE
 volt_tt_606590308c2ef(l_orderkey, l_suppkey) AS (SELECT volt_tt_606590308b512.l_or
    101 | 48237350 | 1073840810 | 12058168 | padb_fetch_sample: select count(*) from
 volt_tt_606590308c2ef
    101 | 48237350 | 1073840810 | 12058170 | SELECT s_name , COUNT(*) AS numwait FROM
 supplier, lineitem l1, orders, nation WHERE s_suppkey = l1.
(5 rows)
```

En la vista de monitoreo SYS SYS_QUERY_HISTORY, Amazon Redshift registra la consulta de la siguiente manera:

```
SELECT user_id, transaction_id, session_id, query_id, query_text::varchar(100)
FROM sys_query_history
WHERE transaction_id = 48237350
ORDER BY start_time;

 user_id | transaction_id | session_id | query_id |
          query_text
---------+----------------+------------+----------
+-----------------------------------------------------------------------------------------------------------
     101 |       48237350 | 1073840810 | 12058149 | SELECT s_name , COUNT(*) AS numwait
 FROM supplier, lineitem l1, orders, nation WHERE s_suppkey = l1.
```

En SYS_QUERY_DETAIL, puede buscar detalles a nivel de consulta secundaria utilizando el valor query_id de SYS_QUERY_HISTORY. La columna child_query_sequence muestra el orden en que se ejecutan las consultas secundarias. Para obtener más información sobre las columnas de SYS_QUERY_DETAIL, consulte SYS_QUERY_DETAIL.

```
select user_id,
```

```
       query_id,
       child_query_sequence,
       stream_id,
       segment_id,
       step_id,
       start_time,
       end_time,
       duration,
       blocks_read,
       blocks_write,
       local_read_io,
       remote_read_io,
       data_skewness,
       time_skewness,
       is_active,
       spilled_block_local_disk,
       spilled_block_remote_disk
from sys_query_detail
where query_id = 12058149
      and step_id = -1
order by query_id,
         child_query_sequence,
         stream_id,
         segment_id,
         step_id;

 user_id | query_id | child_query_sequence | stream_id | segment_id | step_id |
       start_time         |          end_time          | duration | blocks_read |
 blocks_write | local_read_io | remote_read_io | data_skewness | time_skewness |
 is_active | spilled_block_local_disk | spilled_block_remote_disk
---------+----------+----------------------+-----------+------------+---------
+----------------------------+----------------------------+----------+-------------
+--------------+---------------+----------------+---------------+---------------
+-----------+--------------------------+--------------------------
     101 | 12058149 |                    1 |         0 |          0 |      -1 |
 2023-09-27 15:40:38.512415 | 2023-09-27 15:40:38.533333 |    20918 |           0 |
         0 |             0 |              0 |             0 |            44 | f
 |                        0 |                           0
     101 | 12058149 |                    1 |         1 |          1 |      -1 |
 2023-09-27 15:40:39.931437 | 2023-09-27 15:40:39.972826 |    41389 |          12 |
         0 |            12 |              0 |             0 |            77 | f
 |                        0 |                           0
     101 | 12058149 |                    1 |         2 |          2 |      -1 |
 2023-09-27 15:40:40.584412 | 2023-09-27 15:40:40.613982 |    29570 |          32 |
```

```
        0 |            32 |              0 |             0 |            25 | f
|                         0 |                           0
    101 | 12058149 |                    1 |         2 |          3 |      -1 |
2023-09-27 15:40:40.582038 | 2023-09-27 15:40:40.615758 |    33720 |           0 |
        0 |             0 |              0 |             0 |             1 | f
|                         0 |                           0
    101 | 12058149 |                    1 |         3 |          4 |      -1 |
2023-09-27 15:40:46.668766 | 2023-09-27 15:40:46.705456 |    36690 |          24 |
        0 |            15 |              0 |             0 |            17 | f
|                         0 |                           0
    101 | 12058149 |                    1 |         4 |          5 |      -1 |
2023-09-27 15:40:46.707209 | 2023-09-27 15:40:46.709176 |     1967 |           0 |
        0 |             0 |              0 |             0 |            18 | f
|                         0 |                           0
    101 | 12058149 |                    1 |         4 |          6 |      -1 |
2023-09-27 15:40:46.70656  | 2023-09-27 15:40:46.71289  |     6330 |           0 |
        0 |             0 |              0 |             0 |             0 | f
|                         0 |                           0
    101 | 12058149 |                    1 |         5 |          7 |      -1 |
2023-09-27 15:40:46.71405  | 2023-09-27 15:40:46.714343 |      293 |           0 |
        0 |             0 |              0 |             0 |             0 | f
|                         0 |                           0
    101 | 12058149 |                    2 |         0 |          0 |      -1 |
2023-09-27 15:40:52.083907 | 2023-09-27 15:40:52.087854 |     3947 |           0 |
        0 |             0 |              0 |             0 |            35 | f
|                         0 |                           0
    101 | 12058149 |                    2 |         1 |          1 |      -1 |
2023-09-27 15:40:52.089632 | 2023-09-27 15:40:52.091129 |     1497 |           0 |
        0 |             0 |              0 |             0 |            11 | f
|                         0 |                           0
    101 | 12058149 |                    2 |         1 |          2 |      -1 |
2023-09-27 15:40:52.089008 | 2023-09-27 15:40:52.091306 |     2298 |           0 |
        0 |             0 |              0 |             0 |             0 | f
|                         0 |                           0
    101 | 12058149 |                    3 |         0 |          0 |      -1 |
2023-09-27 15:40:56.882013 | 2023-09-27 15:40:56.897282 |    15269 |           0 |
        0 |             0 |              0 |             0 |            29 | f
|                         0 |                           0
    101 | 12058149 |                    3 |         1 |          1 |      -1 |
2023-09-27 15:40:59.718554 | 2023-09-27 15:40:59.722789 |     4235 |           0 |
        0 |             0 |              0 |             0 |            13 | f
|                         0 |                           0
    101 | 12058149 |                    3 |         2 |          2 |      -1 |
2023-09-27 15:40:59.800382 | 2023-09-27 15:40:59.807388 |     7006 |           0 |
```

```
        0 |             0 |              0 |             0 |            58 | f
|                        0 |                           0
    101 | 12058149 |                    3 |         3 |          3 |      -1 |
2023-09-27 15:41:06.488685 | 2023-09-27 15:41:06.493825 |     5140 |           0 |
        0 |             0 |              0 |             0 |            56 | f
|                        0 |                           0
    101 | 12058149 |                    3 |         3 |          4 |      -1 |
2023-09-27 15:41:06.486206 | 2023-09-27 15:41:06.497756 |    11550 |           0 |
        0 |             0 |              0 |             0 |             2 | f
|                        0 |                           0
    101 | 12058149 |                    3 |         4 |          5 |      -1 |
2023-09-27 15:41:06.499201 | 2023-09-27 15:41:06.500851 |     1650 |           0 |
        0 |             0 |              0 |             0 |            15 | f
|                        0 |                           0
    101 | 12058149 |                    3 |         4 |          6 |      -1 |
2023-09-27 15:41:06.498609 | 2023-09-27 15:41:06.500949 |     2340 |           0 |
        0 |             0 |              0 |             0 |             0 | f
|                        0 |                           0
    101 | 12058149 |                    3 |         5 |          7 |      -1 |
2023-09-27 15:41:06.502945 | 2023-09-27 15:41:06.503282 |      337 |           0 |
        0 |             0 |              0 |             0 |             0 | f
|                        0 |                           0
    101 | 12058149 |                    4 |         0 |          0 |      -1 |
2023-09-27 15:41:06.62899  | 2023-09-27 15:41:06.631452 |     2462 |           0 |
        0 |             0 |              0 |             0 |            22 | f
|                        0 |                           0
    101 | 12058149 |                    4 |         1 |          1 |      -1 |
2023-09-27 15:41:06.632313 | 2023-09-27 15:41:06.63391  |     1597 |           0 |
        0 |             0 |              0 |             0 |            20 | f
|                        0 |                           0
    101 | 12058149 |                    4 |         1 |          2 |      -1 |
2023-09-27 15:41:06.631726 | 2023-09-27 15:41:06.633813 |     2087 |           0 |
        0 |             0 |              0 |             0 |             0 | f
|                        0 |                           0
    101 | 12058149 |                    5 |         0 |          0 |      -1 |
2023-09-27 15:41:12.571974 | 2023-09-27 15:41:12.584234 |    12260 |           0 |
        0 |             0 |              0 |             0 |            39 | f
|                        0 |                           0
    101 | 12058149 |                    5 |         0 |          1 |      -1 |
2023-09-27 15:41:12.569815 | 2023-09-27 15:41:12.585391 |    15576 |           0 |
        0 |             0 |              0 |             0 |             4 | f
|                        0 |                           0
    101 | 12058149 |                    5 |         1 |          2 |      -1 |
2023-09-27 15:41:13.758513 | 2023-09-27 15:41:13.76401  |     5497 |           0 |
```

```
         0 |             0 |              0 |             0 |           39 | f
 |                        0 |                          0
     101 | 12058149 |                    5 |         1 |          3 |      -1 |
 2023-09-27 15:41:13.749    | 2023-09-27 15:41:13.772987 |    23987 |           0 |
         0 |             0 |              0 |             0 |           32 | f
 |                        0 |                          0
     101 | 12058149 |                    5 |         2 |          4 |      -1 |
 2023-09-27 15:41:13.799526 | 2023-09-27 15:41:13.813506 |    13980 |           0 |
         0 |             0 |              0 |             0 |           62 | f
 |                        0 |                          0
     101 | 12058149 |                    5 |         2 |          5 |      -1 |
 2023-09-27 15:41:13.798823 | 2023-09-27 15:41:13.813651 |    14828 |           0 |
         0 |             0 |              0 |             0 |            0 | f
 |                        0 |                          0
(28 rows)
```

# Identificadores de consulta, proceso y sesión de tablas del sistema

Al analizar los identificadores de consulta, proceso y sesión que aparecen en las tablas del sistema, tenga en cuenta lo siguiente:

- El valor del identificador de consulta (en columnas como `query_id` y `query`) se puede reutilizar con el tiempo.
- El valor del identificador de proceso o sesión (en columnas como `process_id`, `pid` y `session_id`) se puede reutilizar con el tiempo.
- El valor del identificador de la transacción (en columnas como `transaction_id` y `xid`) es único.

# Vistas de metadatos SVV

Las vistas de SVV son vistas del sistema en Amazon Redshift que contienen información sobre objetos de base de datos. Tenga en cuenta que, aunque almacenan información como los permisos de usuario o los nombres de las tablas, no están pensadas para unirse a relaciones creadas por los usuarios.

> **Note**
>
> Amazon Redshift informa de una ADVERTENCIA, no de un ERROR, si se produce un error en una respuesta de base de datos por cualquier motivo. Amazon Redshift no envía mensajes de ERROR cuando se consultan objetos en un recurso compartido de datos.

Temas


- SVV_ACTIVE_CURSORS
- SVV_ALL_COLUMNS
- SVV_ALL_SCHEMAS
- SVV_ALL_TABLES
- SVV_ALTER_TABLE_RECOMMENDATIONS
- SVV_ATTACHED_MASKING_POLICY
- SVV_COLUMNS
- SVV_COLUMN_PRIVILEGES
- SVV_COPY_JOB_INTEGRATIONS
- SVV_DATABASE_PRIVILEGES
- SVV_DATASHARE_PRIVILEGES
- SVV_DATASHARES
- SVV_DATASHARE_CONSUMERS
- SVV_DATASHARE_OBJECTS
- SVV_DEFAULT_PRIVILEGES
- SVV_DISKUSAGE
- SVV_EXTERNAL_COLUMNS
- SVV_EXTERNAL_DATABASES
- SVV_EXTERNAL_PARTITIONS
- SVV_EXTERNAL_SCHEMAS
- SVV_EXTERNAL_TABLES
- SVV_FUNCTION_PRIVILEGES
- SVV_GEOGRAPHY_COLUMNS

- SVV_GEOMETRY_COLUMNS
- SVV_IAM_PRIVILEGES
- SVV_IDENTITY_PROVIDERS
- SVV_INTEGRATION
- SVV_INTEGRATION_TABLE_MAPPING
- SVV_INTEGRATION_TABLE_STATE
- SVV_INTERLEAVED_COLUMNS
- SVV_LANGUAGE_PRIVILEGES
- SVV_MASKING_POLICY
- SVV_ML_MODEL_INFO
- SVV_ML_MODEL_PRIVILEGES
- SVV_MV_DEPENDENCY
- SVV_MV_INFO
- SVV_QUERY_INFLIGHT
- SVV_QUERY_STATE
- SVV_REDSHIFT_COLUMNS
- SVV_REDSHIFT_DATABASES
- SVV_REDSHIFT_FUNCTIONS
- SVV_REDSHIFT_SCHEMA_QUOTA
- SVV_REDSHIFT_SCHEMAS
- SVV_REDSHIFT_TABLES
- SVV_RELATION_PRIVILEGES
- SVV_RLS_APPLIED_POLICY
- SVV_RLS_ATTACHED_POLICY
- SVV_RLS_POLICY
- SVV_RLS_RELATION
- SVV_ROLE_GRANTS
- SVV_ROLES
- SVV_SCHEMA_PRIVILEGES

- SVV_SCHEMA_QUOTA_STATE
- SVV_SYSTEM_PRIVILEGES
- SVV_TABLE_INFO
- SVV_TABLES
- SVV_TRANSACTIONS
- SVV_USER_GRANTS
- SVV_USER_INFO
- SVV_VACUUM_PROGRESS
- SVV_VACUUM_SUMMARY

# SVV_ACTIVE_CURSORS

SVV_ACTIVE_CURSORS muestra los detalles de los cursores que están abiertos actualmente. Para obtener más información, consulte DECLARE.

SVV_ACTIVE_CURSORS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema. Un usuario solo puede ver los cursores que él mismo haya abierto. Un superusuario puede ver todos los cursores.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | El ID del usuario que creó el cursor. |
| cursor_name | varchar(128) | Nombre del cursor. |
| transaction_id | bigint(128) | El ID de la transacción. |
| session_id | entero | ID del proceso con el cursor activo. |
| declare_time | marca de tiempo | Momento en el que se declaró el cursor. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| total_bytes | bigint | Tamaño del conjunto de resultados del cursor, en bytes. |
| total_rows | bigint | Número de filas del conjunto de resultados del cursor. |
| fetched_rows | bigint | Número de filas que se obtienen actualmente del conjunto de resultados del cursor. |
| cursor_storage_limit_used_percent | entero | Porcentaje de espacio en disco que utiliza actualmente el cursor. |

## SVV_ALL_COLUMNS

Utilice SVV_ALL_COLUMNS para ver una unión de columnas de tablas de Amazon Redshift como se muestra en SVV_REDSHIFT_COLUMNS y la lista consolidada de todas las columnas externas de todas las tablas externas. Para obtener información acerca de las columnas de Amazon Redshift, consulte SVV_REDSHIFT_COLUMNS.

SVV_ALL_COLUMNS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

### Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| database_name | varchar(128) | El nombre de la base de datos. |
| schema_name | varchar(128) | El nombre del esquema. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| table_name | varchar(128) | El nombre de la tabla. |
| column_name | varchar(128) | El nombre de la columna. |
| ordinal_position | entero | La posición de la columna en la tabla. |
| column_default | varchar(4000) | El valor predeterminado de la columna. |
| is_nullable | varchar(3) | Un valor que indica si la columna puede contener valores nulos. Los valores posibles son “yes” (sí) y “no”. |
| data_type | varchar(128) | El tipo de datos de la columna. |
| character_maximum_length | entero | La cantidad máxima de caracteres en la columna. |
| numeric_precision | entero | La precisión numérica. |
| numeric_scale | entero | La escala numérica. |
| remarks | varchar (256) | Observaciones. |

## Consultas de ejemplo

El siguiente ejemplo devuelve el resultado de SVV_ALL_COLUMNS.

```
SELECT *
FROM svv_all_columns
WHERE database_name = 'tickit_db'
    AND TABLE_NAME = 'tickit_sales_redshift'
ORDER BY COLUMN_NAME,
    SCHEMA_NAME
LIMIT 5;
```

```
 database_name | schema_name |     table_name        | column_name | ordinal_position
 | column_default | is_nullable | data_type | character_maximum_length |
 numeric_precision | numeric_scale | remarks
 -------------+------------+---------------------+------------+-----------------
+---------------+------------+----------+-------------------------
+-----------------+--------------+---------
   tickit_db   |    public   | tickit_sales_redshift |    buyerid  |        4         |
               |      NO     |  integer  |                         |         32
  |       0       |
   tickit_db   |    public   | tickit_sales_redshift | commission  |        9         |
               |     YES     |  numeric  |                         |          8
  | 2       |
   tickit_db   |    public   | tickit_sales_redshift |    dateid   |        7         |
               |      NO     |  smallint |                         |         16
  |       0       |
   tickit_db   |    public   | tickit_sales_redshift |   eventid   |        5         |
               |      NO     |  integer  |                         |         32
  |       0       |
   tickit_db   |    public   | tickit_sales_redshift |    listid   |        2         |
               |      NO     |  integer  |                         |         32
  |       0       |
```

# SVV_ALL_SCHEMAS

Utilice SVV_ALL_SCHEMAS para ver una unión de esquemas de Amazon Redshift como se muestra en SVV_REDSHIFT_SCHEMAS y la lista consolidada de todos los esquemas externos de todas las bases de datos. Para obtener más información acerca de esquemas de Amazon Redshift, consulte SVV_REDSHIFT_SCHEMAS.

SVV_ALL_SCHEMAS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| database_name | varchar(128) | El nombre de la base de datos donde se encuentra el esquema. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| schema_name | varchar(128) | El nombre del esquema. |
| schema_owner | entero | El ID de usuario del propietario del esquema. Para obtener información sobre los ID de usuario, consulte PG_USER_INFO. |
| schema_type | varchar(128) | El tipo del esquema. Los valores posibles son esquemas externos, locales y compartidos. |
| schema_acl | varchar(128) | Cadena que define los permisos del usuario o grupo de usuarios especificados para el esquema. |
| source_database | varchar(128) | El nombre de la base de datos de origen para esquemas externos. |
| schema_option | varchar (256) | Las opciones del esquema. Este es un atributo de esquema externo. |

## Consulta de ejemplo

El siguiente ejemplo devuelve el resultado de SVV_ALL_SCHEMAS.

```
SELECT *
FROM svv_all_schemas
WHERE database_name = 'tickit_db'
ORDER BY database_name,
    SCHEMA_NAME;
```

```
 database_name |    schema_name     | schema_owner | schema_type | schema_acl |
 source_database | schema_option
---------------+--------------------+--------------+-------------+------------
+-----------------+---------------
   tickit_db   |       public       |       1      |   shared    |            |
         |
```

# SVV_ALL_TABLES

Utilice SVV_ALL_TABLES para ver una unión de tablas de Amazon Redshift como se muestra en SVV_REDSHIFT_TABLES y la lista consolidada de todas las tablas externas de todos los esquemas externos. Para obtener información acerca de las tablas de Amazon Redshift, consulte SVV_REDSHIFT_TABLES.

SVV_ALL_TABLES es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| database_name | varchar(128) | El nombre de la base de datos donde se encuentra la tabla. |
| schema_name | varchar(128) | Nombre del esquema para la tabla. |
| table_name | varchar(128) | El nombre de la tabla. |
| table_acl | varchar(128) | Cadena que define el permiso del usuario o grupo de usuarios especificados para la tabla. |
| table_type | varchar(128) | El tipo de la tabla. Los valores posibles son vistas, tablas base, tablas externas y tablas compartidas. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| remarks | varchar (256) | Observaciones. |

### Consultas de ejemplo

El siguiente ejemplo devuelve el resultado de SVV_ALL_TABLES.

```
SELECT *
FROM svv_all_tables
WHERE database_name = 'tickit_db'
ORDER BY TABLE_NAME,
    SCHEMA_NAME
LIMIT 5;

 database_name | schema_name |        table_name        | table_type | table_acl |
 remarks
---------------+-------------+--------------------------+------------+-----------
+---------
   tickit_db   |    public   | tickit_category_redshift |    TABLE   |           |
   tickit_db   |    public   |   tickit_date_redshift   |    TABLE   |           |
   tickit_db   |    public   |   tickit_event_redshift  |    TABLE   |           |
   tickit_db   |    public   | tickit_listing_redshift  |    TABLE   |           |
   tickit_db   |    public   |   tickit_sales_redshift  |    TABLE   |           |
```

Si el valor de table_acl es nulo, no se han concedido explícitamente privilegios de acceso a la tabla correspondiente.

## SVV_ALTER_TABLE_RECOMMENDATIONS

Registra las recomendaciones actuales de Amazon Redshift Advisor para tablas. En esta vista se muestran las recomendaciones para todas las tablas, tanto si están definidas para la optimización automática como si no. Para ver si una tabla está definida para la optimización automática, consulte SVV_TABLE_INFO. Las entradas aparecen únicamente para las tablas visibles en la base de datos de la sesión actual. Una vez aplicada una recomendación, sea por Amazon Redshift o por usted mismo, ya no aparecerá en la vista.

Solo los superusuarios pueden ver SVV_ALTER_TABLE_RECOMMENDATIONS. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| type | character (30) | El tipo de recomendación. Los valores posibles son distkey y sortkey. |
| database | character (128) | El nombre de la base de datos. |
| table_id | entero | El identificador de tabla. |
| group_id | entero | El número de grupo de un conjunto de recomendaciones. Todas las recomendaciones de un grupo deben aplicarse para ver el máximo beneficio. Los valores posibles son -1 para una recomendación de clave de ordenación y un número mayor que cero para una recomendación de clave de distribución. |
| ddl | character (1024) | Instrucción SQL que se debe ejecutar para aplicar la recomendación. |
| auto_eligible | character (1) | El valor indica si la recomendación es apta para que Amazon Redshift se ejecute automáticamente. Si este valor es `t`, entonces la indicación es true (verdadera); si es `f`, es false (falsa). |

## Consultas de ejemplo

En el siguiente ejemplo, las filas del resultado muestran recomendaciones para la clave de distribución y la clave de ordenación. Las filas también muestran si las recomendaciones son aptas para que Amazon Redshift las aplique automáticamente.

```
select type, database, table_id, group_id, ddl, auto_eligible
from svv_alter_table_recommendations;
```

```
type      | database | table_id | group_id | ddl

                  | auto_eligible
diststyle | db0      | 117884   | 2        | ALTER TABLE "sch"."dp21235_tbl_1" ALTER
DISTSTYLE KEY DISTKEY "c0"
                    | f
diststyle | db0      | 117892   | 2        | ALTER TABLE "sch"."dp21235_tbl_1" ALTER
DISTSTYLE KEY DISTKEY "c0"
                    | f
diststyle | db0      | 117885   | 1        | ALTER TABLE "sch"."catalog_returns"
ALTER DISTSTYLE KEY DISTKEY "cr_sold_date_sk", ALTER COMPOUND SORTKEY
("cr_sold_date_sk","cr_returned_time_sk") | t
sortkey   | db0      | 117890   | -1       | ALTER TABLE "sch"."customer_addresses"
ALTER COMPOUND SORTKEY ("ca_address_sk")
                     | t
```

# SVV_ATTACHED_MASKING_POLICY

Utilice SVV_ATTACHED_MASKING_POLICY para ver todas las relaciones y los roles o los usuarios con políticas adjuntas a la base de datos conectada actualmente.

Solo los superusuarios y los usuarios con el rol `sys:secadmin` pueden ver SVV_ATTACHED_MASKING_POLICY. Los usuarios normales verán 0 filas.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| policy_name | texto | Nombre de la política de enmascaramiento adjunta a la tabla. |
| schema_name | texto | Esquema de la tabla a la que está asociada la política. |
| table_name | texto | Nombre de la tabla a la que está asociada la política. |
| table_type | texto | Tipo de la tabla a la que está asociada la política. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| grantor | texto | Nombre del usuario que asoció la política. |
| grantee | texto | Nombre del usuario o rol al que está asociada la política. |
| grantee_type | texto | El tipo de beneficiario. Puede ser role, user o public. |
| priority | int | Prioridad de la política adjunta. |
| input_columns | texto | Atributos de la columna de entrada de la política adjunta. |
| output_columns | texto | Atributos de la columna de salida de la política adjunta. |
| is_masking_datashare_on | booleano | Si la tabla a la que está asociada la política está protegida por DDM a través de recursos compartidos de datos. |

## Funciones internas

SVV_ATTACHED_MASKING_POLICY admite las siguientes funciones internas:

mask_get_policy_for_role_on_column

Obtiene la política de máxima prioridad que se aplica a un determinado par de columna-rol.

Sintaxis

```
mask_get_policy_for_role_on_column
                        (relschema,
                        relname,
                        colname,
```

```
                        rolename);
```

Parámetros

relschema

Nombre del esquema en el que se encuentra la política.

relname

Nombre de la tabla en la que se encuentra la política.

colname

Nombre de la columna a la que está asociada la política.

rolename

Nombre del rol al que está asociada la política.

mask_get_policy_for_user_on_column

Obtiene la política de mayor prioridad que se aplica a un determinado par de columna-usuario.

Sintaxis

```
mask_get_policy_for_user_on_column
                        (relschema,
                        relname,
                        colname,
                        username);
```

Parámetros

relschema

Nombre del esquema en el que se encuentra la política.

relname

Nombre de la tabla en la que se encuentra la política.

colname

Nombre de la columna a la que está asociada la política.

rolename

Nombre del usuario al que está asociada la política.

## SVV_COLUMNS

Utilice SVV_COLUMNS para consultar información del catálogo sobre las columnas de las vistas y tablas tanto locales como externas, incluidas las vistas de enlace de tiempo de ejecución.

SVV_COLUMNS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

La vista SVV_COLUMNS combina metadatos de la tabla de las vistas del sistema Tablas de catálogos de sistema (tablas con un prefijo PG) y SVV_EXTERNAL_COLUMNS. Las tablas del catálogo del sistema describen las tablas de base de datos de Amazon Redshift. SVV_EXTERNAL_COLUMNS describe las tablas externas que se utilizan con Amazon Redshift Spectrum.

Cualquier usuario puede ver todas las filas de las tablas de catálogo del sistema. Los usuarios normales pueden ver las definiciones de las columnas de la vista SVV_EXTERNAL_COLUMNS únicamente para las tablas externas para las que se les ha concedido acceso. Aunque los usuarios normales pueden ver los metadatos de las tablas de catálogo del sistema, solo pueden seleccionar datos de las tablas definidas por el usuario si son propietarios de la tabla o se les ha concedido acceso.

### Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| table_catalog | texto | Nombre del catálogo donde está la tabla. |
| table_schema | texto | Nombre del esquema para la tabla. |
| table_name | texto | El nombre de la tabla. |
| column_name | texto | El nombre de la columna. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| ordinal_position | int | La posición de la columna en la tabla. |
| column_default | texto | El valor predeterminado de la columna. |
| is_nullable | texto | Un valor que indica si la columna puede contener valores nulos. |
| data_type | texto | El tipo de datos de la columna. |
| character_maximum_length | int | La cantidad máxima de caracteres en la columna. |
| numeric_precision | int | La precisión numérica. Si la columna data_type es numérica, devuelve el número de dígitos significativos del valor entero. |
| numeric_precision_radix | int | La base de precisión numérica. Si la columna data_type es numérica, devuelve la base de las columnas numeric_precision y numeric_scale. |
| numeric_scale | int | La escala numérica. Si la columna data_type es numérica, devuelve el número de dígitos significativos del valor decimal. |
| datetime_precision | int | La precisión de fecha y hora. |
| interval_type | texto | El tipo de intervalo. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| interval_precision | texto | La precisión del intervalo. |
| character_set_catalog | texto | El catálogo del conjunto de caracteres. |
| character_set_schema | texto | El esquema del conjunto de caracteres. |
| character_set_name | texto | El nombre del conjunto de caracteres. |
| collation_catalog | texto | El catálogo de intercalación. |
| collation_schema | texto | El esquema de la intercala ción. |
| collation_name | texto | El nombre de la intercalación. |
| domain_name | texto | El nombre del dominio. |
| remarks | texto | Observaciones. |

## SVV_COLUMN_PRIVILEGES

Utilice SVV_COLUMN_PRIVILEGES para ver los permisos de columna que tienen concedidos de forma explícita los usuarios, roles y grupos de la base de datos actual.

SVV_COLUMN_PRIVILEGES es visible para los siguientes usuarios:

- Superusuarios
- Usuarios con el permiso ACCESS SYSTEM TABLE

Los demás usuarios solo pueden ver las identidades a las que tienen acceso o que son de su propiedad.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| namespace_name | texto | Nombre del espacio de nombres donde existe una relación especific ada. |
| relation_name | texto | Nombre de la relación. |
| column_name | texto | El nombre de la columna. |
| privilege_type | texto | Tipo de permiso. Los valores posibles son SELECT o UPDATE. |
| identity_id | entero | ID de la identidad. Los valores posibles son ID de usuario, ID de rol o ID de grupo. |
| identity_name | texto | Nombre de la identidad. |
| identity_type | texto | El tipo de la identidad. Los valores posibles son usuario, rol, grupo o público. |

## Consulta de ejemplo

En el siguiente ejemplo, se muestra el resultado de SVV_COLUMN_PRIVILEGES.

```
SELECT
 namespace_name,relation_name,COLUMN_NAME,privilege_type,identity_name,identity_type
FROM svv_column_privileges WHERE relation_name = 'lineitem';

 namespace_name | relation_name | column_name | privilege_type | identity_name |
 identity_type
----------------+---------------+-------------+----------------+---------------
+----------------
    public      |   lineitem    | l_orderkey  |     SELECT     |    reguser    |
 user
```

```
   public      |   lineitem    | l_orderkey  |     SELECT     |     role1     |
role
   public      |   lineitem    | l_partkey   |     SELECT     |    reguser    |
user
   public      |   lineitem    | l_partkey   |     SELECT     |     role1     |
role
```

# SVV_COPY_JOB_INTEGRATIONS

Utilice SVV_COPY_JOB_INTEGRATIONS para ver los detalles de las integraciones de eventos de S3.

Esta vista contiene las integraciones de eventos de S3 que se han creado.

SVV_COPY_JOB_INTEGRATIONS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| job_owner | entero | Identificador del propietario del trabajo. |
| channel_arn | nombre | Identificador de la integración. |
| bucket | texto | Nombre del bucket de Amazon S3 asociado con la integración. |
| channel state | nombre | Estado de la integración. Valores aceptados: `Pending`, `Established` y `Inactive`. |
| db_name | nombre | Nombre de la base de datos del objeto dependiente. |
| job_name | texto | El nombre del trabajo. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| job_state | entero | Estado del trabajo. Valores válidos: 0 para activos, 1 para pendientes. |

El siguiente ejemplo devuelve las integraciones de S3 para la base de datos actual.

```
SELECT * FROM SVV_COPY_JOB_INTEGRATIONS WHERE db_name = pg_catalog.current_database();
```

# SVV_DATABASE_PRIVILEGES

Utilice SVV_DATABASE_PRIVILEGES para ver los permisos de base de datos que tienen concedidos de forma explícita los usuarios, roles y grupos del clúster de Amazon Redshift.

SVV_DATABASE_PRIVILEGES es visible para los siguientes usuarios:

- Superusuarios
- Usuarios con el permiso ACCESS SYSTEM TABLE

Los demás usuarios solo pueden ver las identidades a las que tienen acceso o que son de su propiedad.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| database_name | texto | El nombre de la base de datos. |
| privilege_type | texto | Tipo de permiso. Para los permisos con privilege_scope de base de datos, los valores posibles son USAGE, CREATE, TEMPORARY, TEMP y ALTER. Para valores de privilege_scope distintos de la base de datos, los valores posibles incluyen cualquier tipo de permiso disponible en el ámbito del permiso. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| identity_id | entero | ID de la identidad. Los valores posibles son ID de usuario, ID de rol o ID de grupo. |
| identity_name | texto | Nombre de la identidad. |
| identity_type | texto | El tipo de la identidad. Los valores posibles son usuario, rol, grupo o público. |
| admin_option | booleano | Valor que indica si el usuario puede conceder el permiso a otros usuarios y roles. Siempre es false (falso) para el tipo de identidad de rol y grupo. |
| privilege_scope | texto | El alcance del permiso especificado en privilege_type. Los valores posibles son los siguientes:<br>• DATABASE<br>• SCHEMAS<br>• TABLES<br>• FUNCTIONS<br>• LANGUAGES<br><br>Para obtener información sobre los permisos acotados, vaya a Permisos acotados. |

## Consulta de ejemplo

En el siguiente ejemplo, se muestra el resultado de SVV_DATABASE_PRIVILEGES.

```
SELECT database_name,privilege_type,identity_name,identity_type,admin_option FROM
 svv_database_privileges
WHERE database_name = 'test_db';
```

```
 database_name | privilege_type | identity_name | identity_type | admin_option
---------------+----------------+---------------+---------------+--------------
     test_db   |     CREATE     |     reguser   |      user     |     False
     test_db   |     CREATE     |      role1    |      role     |     False
     test_db   |     TEMP       |      public   |      public   |     False
     test_db   |     TEMP       |      role1    |      role     |     False
```

# SVV_DATASHARE_PRIVILEGES

Utilice SVV_DATASHARE_PRIVILEGES para ver los permisos de recurso compartido de datos que tienen concedidos de forma explícita los usuarios, roles y grupos del clúster de Amazon Redshift.

SVV_DATASHARE_PRIVILEGES es visible para los siguientes usuarios:

- Superusuarios
- Usuarios con el permiso ACCESS SYSTEM TABLE

Los demás usuarios solo pueden ver las identidades a las que tienen acceso o que son de su propiedad.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| datashare_name | texto | El nombre del datashare. |
| privilege_type | texto | Tipo de permiso. Los valores posibles son ALTER o SHARE. |
| identity_id | entero | ID de la identidad. Los valores posibles son ID de usuario, ID de rol o ID de grupo. |
| identity_name | texto | Nombre de la identidad. |
| identity_type | texto | El tipo de la identidad. Los valores posibles son usuario, rol, grupo o público. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| admin_option | booleano | Valor que indica si el usuario puede conceder el permiso a otros usuarios y roles. Siempre es false (falso) para el tipo de identidad de rol y grupo. |

### Consulta de ejemplo

En el siguiente ejemplo, se muestra el resultado de SVV_DATASHARE_PRIVILEGES.

```
SELECT datashare_name,privilege_type,identity_name,identity_type,admin_option FROM
 svv_datashare_privileges
WHERE datashare_name = 'demo_share';

 datashare_name | privilege_type |  identity_name | identity_type | admin_option
----------------+----------------+----------------+---------------+--------------
   demo_share   |     ALTER      |    superuser   |     user      |   False
   demo_share   |     ALTER      |    reguser     |     user      |   False
```

## SVV_DATASHARES

Utilice SVV_DATASHARES para ver una lista de los recursos compartidos de datos creados en el clúster, así como de los que están compartidos con el clúster.

SVV_DATASHARES es visible para los siguientes usuarios:

- Superusuarios
- Propietarios del recurso compartido de datos
- Usuarios con permisos ALTER o USAGE en un recurso compartido de datos

Otros usuarios no pueden ver ninguna fila. Para obtener más información sobre los permisos ALTER y USAGE, consulte GRANT.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| share_name | varchar(128) | El nombre de un datashare. |
| share_id | entero | El ID del datashare. |
| share_owner | entero | El propietario del datashare. |
| source_database | varchar(128) | La base de datos de origen para este datashare. |
| consumer_database | varchar(128) | La base de datos consumido ra que se crea a partir de este datashare. |
| share_type | varchar(8) | El tipo de datashare. Los valores posibles son INBOUND y OUTBOUND. |
| createdate | TIMESTAMP sin zona horaria | La fecha de creación del datashare. |
| is_publicaccessible | booleano | La propiedad que especific a si un datashare se puede compartir en un clúster de acceso público. |
| share_acl | varchar (256) | Cadena que define los permisos del usuario o grupo de usuarios especificados para el recurso compartido de datos. |
| producer_account | varchar (16) | El ID de la cuenta productora del datashare. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| producer_namespace | varchar (64) | El identificador de clúster único para el clúster productor del datashare. |
| managed_by | varchar (64) | La propiedad que especifica el servicio de AWS que administra el recurso compartido de datos. |

## Notas de uso

Recuperación de metadatos adicionales: con el número entero devuelto en la columna `share_owner`, puede unirlo con `usesysid` en SVL_USER_INFO para obtener datos sobre el propietario del recurso compartido de datos. Esto incluye el nombre y las propiedades adicionales.

## Consulta de ejemplo

El siguiente ejemplo devuelve la salida para SVV_DATASHARES.

```
SELECT share_owner, source_database, share_type, is_publicaccessible
FROM svv_datashares
WHERE share_name LIKE 'tickit_datashare%'
AND source_database = 'dev';

  share_owner | source_database | share_type  | is_publicaccessible
--------------+-----------------+-------------+---------------------
     100      |      dev        |   OUTBOUND  |        True
(1 rows)
```

El siguiente ejemplo devuelve el resultado para SVV_DATASHARES para los recursos compartidos de datos de salida.

```
SELECT share_name, share_owner, btrim(source_database), btrim(consumer_database),
 share_type, is_publicaccessible, share_acl, btrim(producer_account),
 btrim(producer_namespace), btrim(managed_by) FROM svv_datashares WHERE share_type =
 'OUTBOUND';
```

```
   share_name   | share_owner | source_database | consumer_database | share_type |
 is_publicaccessible | share_acl | producer_account|         producer_namespace
   | managed_by
---------------+-------------+-----------------+-------------------
+------------+---------------------+-----------+-----------------
+--------------------------------------+------------
    salesshare  |      1      |       dev       |                   |  OUTBOUND  |
   True         |           |   123456789012  | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d |

 marketingshare |      1      |       dev       |                   |  OUTBOUND  |
   True         |           |   123456789012  | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d |
```

El siguiente ejemplo devuelve el resultado para SVV_DATASHARES para los recursos compartidos de datos de entrada.

```
SELECT share_name, share_owner, btrim(source_database), btrim(consumer_database),
 share_type, is_publicaccessible, share_acl, btrim(producer_account),
 btrim(producer_namespace), btrim(managed_by) FROM svv_datashares WHERE share_type =
 'INBOUND';

  share_name    | share_owner | source_database | consumer_database | share_type |
 is_publicaccessible | share_acl | producer_account |         producer_namespace
    | managed_by
---------------+-------------+-----------------+-------------------
+------------+---------------------+-----------+-----------------
+--------------------------------------+------------
  salesshare    |             |                 |                   |  INBOUND   |
   False         |           |  123456789012    | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d
 |
 marketingshare |             |                 |                   |  INBOUND   |
  False         |           |  123456789012    | 13b8833d-17c6-4f16-8fe4-1a018f5ed00d |
 ADX
```

# SVV_DATASHARE_CONSUMERS

Utilice SVV_DATASHARE_CONSUMERS para ver una lista de consumidores de un datashare creado en un clúster.

SVV_DATASHARE_CONSUMERS es visible para los siguientes usuarios:

- Superusuarios

- Propietarios del recurso compartido de datos
- Usuarios con permisos ALTER o USAGE en un recurso compartido de datos

Otros usuarios no pueden ver ninguna fila. Para obtener más información sobre los permisos ALTER y USAGE, consulte GRANT.

### Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| share_name | varchar(128) | El nombre del datashare. |
| consumer_account | varchar (16) | El ID de cuenta para el consumidor del datashare. |
| consumer_namespace | varchar (64) | El identificador de clúster único del clúster consumidor del datashare. |
| share_date | TIMESTAMP sin zona horaria | La fecha en que se compartió el datashare. |

### Consulta de ejemplo

El siguiente ejemplo devuelve el resultado de SVV_DATASHARE_CONSUMERS.

```
SELECT count(*)
FROM svv_datashare_consumers
WHERE share_name LIKE 'tickit_datashare%';

1
```

## SVV_DATASHARE_OBJECTS

Utilice SVV_DATASHARE_OBJECTS para ver una lista de objetos de todos los datashares creados en el clúster o compartidos con el clúster.

SVV_DATASHARE_OBJECTS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Para obtener información sobre cómo ver una lista de recursos compartidos de datos, consulte SVV_DATASHARES.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| share_type | varchar(8) | El tipo del datashare especificado. Los valores posibles son OUTBOUND e INBOUND. |
| share_name | varchar(128) | El nombre del datashare. |
| object_type | varchar (64) | El tipo de un objeto especificado. Los valores posibles son esquemas, tablas, vistas, vistas de enlace de tiempo de ejecución, vistas materializadas y funciones. |
| object_name | varchar(512) | El nombre del objeto. El nombre del objeto se extiende para incluir el nombre del esquema, como schema1.t1. |
| producer_account | varchar (16) | El ID de la cuenta productora del datashare. |
| producer_namespace | varchar (64) | El identificador de clúster único para el clúster productor del datashare. |
| include_new | booleano | La propiedad que especifica si se agregarán futuras tablas, vistas o funciones |

| Nombre de la columna | Tipo de datos: | Descripción |
| --- | --- | --- |
| | | SQL definidas por el usuario (UDF) creadas en el esquema especificado al datashare . Este parámetro solo es relevante para los datashare s OUTBOUND y solo para los tipos de esquema en el datashare. |

## Consulta de ejemplo

Los siguientes ejemplos devuelven el resultado de SVV_DATASHARE_OBJECTS.

```
SELECT share_type,
    btrim(share_name)::varchar(16) AS share_name,
    object_type,
    object_name
FROM svv_datashare_objects
WHERE share_name LIKE 'tickit_datashare%'
AND object_name LIKE '%tickit%'
ORDER BY object_name
LIMIT 5;

 share_type |     share_name     | object_type |          object_name
------------+--------------------+-------------+--------------------------------
 OUTBOUND   |  tickit_datashare  |    table    |  public.tickit_category_redshift
 OUTBOUND   |  tickit_datashare  |    table    |  public.tickit_date_redshift
 OUTBOUND   |  tickit_datashare  |    table    |  public.tickit_event_redshift
 OUTBOUND   |  tickit_datashare  |    table    |  public.tickit_listing_redshift
 OUTBOUND   |  tickit_datashare  |    table    |  public.tickit_sales_redshift
```

```
SELECT * FROM SVV_DATASHARE_OBJECTS WHERE share_name like 'sales%';

share_type | share_name | object_type | object_name  | producer_account |
 producer_namespace          | include_new
-----------+------------+-------------+--------------+------------------
+--------------------------------------+-------------
```

```
OUTBOUND  | salesshare | schema      | public       | 123456789012     |
13b8833d-17c6-4f16-8fe4-1a018f5ed00d |      t
OUTBOUND  | salesshare | table       | public.sales | 123456789012     |
13b8833d-17c6-4f16-8fe4-1a018f5ed00d |
```

# SVV_DEFAULT_PRIVILEGES

Utilice SVV_DEFAULT_PRIVILEGES para ver los privilegios predeterminados a los que tiene acceso un usuario en un clúster de Amazon Redshift.

SVV_DEFAULT_PRIVILEGES es visible para los siguientes usuarios:

- Superusuarios
- Usuarios con el permiso ACCESS SYSTEM TABLE

Los demás usuarios solo pueden ver los permisos predeterminados que se les concedan.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| schema_name | texto | El nombre del esquema. |
| object_type | texto | El tipo de objeto. Los valores posibles son RELATION, FUNCTION o PROCEDURE. |
| owner_id | entero | El ID del propietario. El valor posible es el ID de usuario. |
| owner_name | texto | El nombre del propietario. |
| owner_type | texto | El tipo de propietario. El valor posible es el usuario. |
| privilege_type | texto | El tipo de privilegio. Los valores posibles son INSERT, SELECT, UPDATE, DELETE, RULE, REFERENCES TRIGGER, DROP y EXECUTE. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| grantee_id | entero | El ID del beneficiario de la concesión. Los valores posibles son ID de usuario, ID de rol e ID de grupo. |
| grantee_type | texto | El tipo de beneficiario de la concesión. Los valores posibles son usuario, rol y público. |
| admin_option | booleano | Este valor que indica si el usuario puede conceder permisos a otros usuarios y roles. Siempre es false (falso) para el tipo de rol y grupo. |

### Consulta de ejemplo

En el siguiente ejemplo se devuelve la salida para SVV_DEFAULT_PRIVILEGES.

```
SELECT * from svv_default_privileges;

 schema_name |    object_type    | owner_id | owner_name | owner_type | privilege_type
 | grantee_id | grantee_name | grantee_type | admin_option
-------------+-------------------+--------- +------------+------------+---------------
+------------+--------------+--------------+-------------+
   public    |     RELATION      |    106   |     u1     |    user    |     UPDATE
 |     107    |      u2      |     user     |      f      |
   public    |     RELATION      |    106   |     u1     |    user    |     SELECT
 |     107    |      u2      |     user     |      f      |
```

## SVV_DISKUSAGE

Amazon Redshift crea la vista de sistema SVV_DISKUSAGE mediante la combinación de las tablas STV_TBL_PERM y STV_BLOCKLIST. La vista SVV_DISKUSAGE tiene información relacionada con la asignación de datos para las tablas en una base de datos.

Utilice consultas de agregación con SVV_DISKUSAGE, como se muestra en los siguientes ejemplos, para determinar la cantidad de bloques de disco asignados a cada base de datos, tabla, sector o columna. Cada bloque de datos utiliza 1 MB. También puede utilizar STV_PARTITIONS para obtener información resumida acerca de la utilización del disco.

Solo los superusuarios pueden ver SVV_DISKUSAGE. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> Esta vista solo está disponible cuando se consultan clústeres aprovisionados.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| db_id | entero | ID de la base de datos. |
| nombre | character (72) | Nombre de la tabla. |
| slice | entero | Sector de datos asignado a la tabla. |
| col | entero | Índice con base cero para la columna. Cada tabla que cree tiene tres columnas ocultas anexadas: INSERT_XID, DELETE_XID y ROW_ID (OID). Una tabla con 3 columnas definidas por el usuario tiene 6 columnas reales, y las columnas definidas por el usuario se enumeran internamente como 0, 1 y 2. Las columnas INSERT_XID, DELETE_XID y ROW_ID se enumeran 3, 4 y 5, respectivamente, en este ejemplo. |
| tbl | entero | ID de la tabla. |
| blocknum | entero | ID para el bloque de datos. |
| num_values | entero | Cantidad de valores contenidos en el bloque. |
| minvalue | bigint | Valor mínimo contenido en el bloque. |
| maxvalue | bigint | Valor máximo contenido en el bloque. |
| sb_pos | entero | Identificador interno para la posición del super bloque en el disco. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| pinned | entero | Indica si el bloque se conectó o no a la memoria como parte de la carga previa: 0 = false; 1 = true. La opción predeterminada es falso. |
| on_disk | entero | Indica si el bloque se almacenó automáticamente o no en el disco: 0 = false; 1 = true. La opción predeterminada es falso. |
| modified | entero | Indica si el bloque se modificó o no: 0 = false; 1 = true. La opción predeterminada es falso. |
| hdr_modified | entero | Indica si el encabezado del bloque se modificó o no: 0 = false; 1 = true. La opción predeterminada es falso. |
| unsorted | entero | Indica si el bloque está desordenado o no: 0 = false; 1 = true. El valor predeterminado es verdadero. |
| tombstone | entero | Para uso interno. |
| preferred _diskno | entero | Cantidad de discos en que debe estar el bloque, excepto que el disco tenga una falla. Una vez arreglado el disco, el bloque volverá a ese disco. |
| temporary | entero | Indica si el bloque tiene o no datos temporales, como una tabla temporal o resultados intermedios de consulta: 0 = false; 1 = true. La opción predeterminada es falso. |
| newblock | entero | Indica si un bloque es o no nuevo (true) o si nunca se guardó en el disco (false): 0 = false; 1 = true. |

## Consultas de ejemplo

SVV_DISKUSAGE tiene una fila por cada bloque del disco designado, por lo que una consulta que selecciona todas las filas posiblemente devuelva una gran cantidad de filas. Le recomendamos usar solo las consultas de agregación con SVV_DISKUSAGE.

Devuelve la mayor cantidad de bloques que alguna vez se asignaron a la columna 6 en la tabla USERS (la columna EMAIL):

```
select db_id, trim(name) as tablename, max(blocknum)
from svv_diskusage
where name='users' and col=6
group by db_id, name;

db_id  | tablename | max
--------+-----------+-----
175857 | users     |   2
(1 row)
```

La siguiente consulta devuelve resultados similares para todas las columnas en una tabla grande de 10 columnas llamada SALESNEW. (Las últimas tres filas, para las columnas de la 10 a la 12, son para las columnas ocultas de metadatos).

```
select db_id, trim(name) as tablename, col, tbl, max(blocknum)
from svv_diskusage
where name='salesnew'
group by db_id, name, col, tbl
order by db_id, name, col, tbl;

db_id  | tablename  | col |  tbl   | max
--------+------------+-----+--------+-----
175857 | salesnew   |   0 | 187605 | 154
175857 | salesnew   |   1 | 187605 | 154
175857 | salesnew   |   2 | 187605 | 154
175857 | salesnew   |   3 | 187605 | 154
175857 | salesnew   |   4 | 187605 | 154
175857 | salesnew   |   5 | 187605 |  79
175857 | salesnew   |   6 | 187605 |  79
175857 | salesnew   |   7 | 187605 | 302
175857 | salesnew   |   8 | 187605 | 302
175857 | salesnew   |   9 | 187605 | 302
175857 | salesnew   |  10 | 187605 |   3
175857 | salesnew   |  11 | 187605 |   2
175857 | salesnew   |  12 | 187605 | 296
(13 rows)
```

# SVV_EXTERNAL_COLUMNS

Utilice SVV_EXTERNAL_COLUMNS para ver los detalles de las columnas en las tablas externas. Además, utilice SVV_EXTERNAL_COLUMNS para consultas entre bases de datos con el fin de ver

detalles de todas las columnas de la tabla en bases de datos no conectadas a las que los usuarios tienen acceso.

SVV_EXTERNAL_COLUMNS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| redshift_database_name | texto | Nombre de la base de datos de Amazon Redshift. |
| schemaname | texto | El nombre del esquema externo de Amazon Redshift para la tabla externa. |
| tablename | texto | Nombre de la tabla externa. |
| columnname | texto | El nombre de la columna. |
| external_type | texto | El tipo de datos de la columna. |
| columnnum | entero | El número de columna externa, a partir de 1. |
| part_key | entero | Si la columna es una clave de partición, el orden de la clave. Si la columna no es una partición, el valor es **0**. |
| is_nullable | texto | Define si una columna se puede anular o no. Algunos valores son `true`, `false` o " " cadena vacía que no representa ninguna informaci ón. |

# SVV_EXTERNAL_DATABASES

Utilice SVV_EXTERNAL_DATABASES para ver los detalles de las bases de datos externas.

SVV_EXTERNAL_DATABASES es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| eskind | entero | El tipo de catálogo externo de la base de datos **1** indica un catálogo de datos; **2** indica un metastore de Hive. |
| esoptions | texto | Detalles del catálogo donde reside la base de datos. |
| databasename | texto | Nombre de la base de datos donde está el catálogo externo |
| ubicación | texto | Ubicación de la base de datos. |
| parameters | texto | Parámetros de la base de datos. |

# SVV_EXTERNAL_PARTITIONS

Utilice SVV_EXTERNAL_PARTITIONS para ver los detalles de las particiones en las tablas externas.

SVV_EXTERNAL_PARTITIONS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

### Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| schemaname | texto | El nombre del esquema externo de Amazon Redshift para la tabla externa con las particiones especificadas. |
| tablename | texto | Nombre de la tabla externa. |
| valores | texto | Valores para la partición. |
| ubicación | texto | Ubicación de la partición. El tamaño de columna está limitado a 128 caracteres. Los valores más largos se truncan. |
| input_format | texto | El formato de entrada. |
| output_format | texto | El formato de salida. |
| serialization_lib | texto | La biblioteca de serialización. |
| serde_parameters | texto | Parámetros SerDe. |
| compressed | entero | Valor que indica si la partición está comprimida: **1** indica que está comprimida; **0** indica que no lo está. |
| parameters | texto | Propiedades de la partición. |

## SVV_EXTERNAL_SCHEMAS

Utilice SVV_EXTERNAL_SCHEMAS para ver información acerca de los esquemas externos. Para obtener más información, consulte CREATE EXTERNAL SCHEMA.

SVV_EXTERNAL_SCHEMAS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| esoid | oid | ID del esquema externo. |
| eskind | smallint | El tipo de catálogo externo para el esquema externo: 1 indica un catálogo de datos, 2 indica un metastore Hive, 3 indica una consulta federada a Aurora PostgreSQL o Amazon RDS PostgreSQL, 4 indica un esquema para una base de datos local de Amazon Redshift, 5 indica un esquema para una base de datos remota de Amazon Redshift, 6 indica un esquema para una tabla del sistema, 8 indica un esquema para bases de datos MySQL remotas, 9 indica un esquema para un flujo de datos de Amazon Kinesis y 10 indica un flujo de datos de Amazon Managed Streaming para Apache Kafka. |
| schemaname | nombre | Nombre de esquema externo. |
| esowner | entero | ID de usuario del propietario del esquema externo. |
| databasename | texto | Nombre de base de datos externa. |
| esoptions | texto | Opciones de esquemas externos. |

## Ejemplo

En el siguiente ejemplo, se muestran detalles de esquemas externos.

```
select * from svv_external_schemas;

esoid  | eskind | schemaname | esowner | databasename | esoptions

-------+--------+------------+---------+--------------
+-------------------------------------------------------------
```

```
100133 |      1 | spectrum   |     100 | redshift     |
 {"IAM_ROLE":"arn:aws:iam::123456789012:role/mySpectrumRole"}
```

# SVV_EXTERNAL_TABLES

Utilice SVV_EXTERNAL_TABLES para ver los detalles de las tablas externas. Para obtener más información, consulte CREATE EXTERNAL SCHEMA. Además, utilice SVV_EXTERNAL_TABLES para consultas entre bases de datos con el fin de ver metadatos en todas las tablas de bases de datos no conectadas a las que los usuarios tienen acceso.

SVV_EXTERNAL_TABLES es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| redshift_database_name | texto | Nombre de la base de datos de Amazon Redshift. |
| schemaname | texto | El nombre del esquema externo de Amazon Redshift para la tabla externa. |
| tablename | texto | Nombre de la tabla externa. |
| tabletype | texto | El tipo de tabla. Algunos valores son TABLE, VIEW, MATERIALIZED VIEW o cadena vacía " " que no representa ninguna informaci ón. |
| ubicación | texto | La ubicación de la tabla. |
| input_format | texto | El formato de entrada |
| output_format | texto | El formato de salida. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| serialization_lib | texto | La biblioteca de serialización. |
| serde_parameters | texto | Parámetros SerDe. |
| compressed | entero | Valor que indica si la tabla está comprimida: **1** indica que está comprimida; **0** indica que no lo está. |
| parameters | texto | Propiedades de tabla. |

### Ejemplo

El siguiente ejemplo muestra detalles svv_external_tables con un predicado en el esquema externo utilizado por una consulta federada.

```
select schemaname, tablename from svv_external_tables where schemaname = 'apg_tpch';
schemaname  | tablename
------------+-----------
apg_tpch    | customer
apg_tpch    | lineitem
apg_tpch    | nation
apg_tpch    | orders
apg_tpch    | part
apg_tpch    | partsupp
apg_tpch    | region
apg_tpch    | supplier
(8 rows)
```

## SVV_FUNCTION_PRIVILEGES

Utilice SVV_FUNCTION_PRIVILEGES para ver los permisos de función que tienen concedidos de forma explícita los usuarios, roles y grupos de la base de datos actual.

SVV_FUNCTION_PRIVILEGES es visible para los siguientes usuarios:

- Superusuarios
- Usuarios con el permiso ACCESS SYSTEM TABLE

Los demás usuarios solo pueden ver las identidades a las que tienen acceso o que son de su propiedad.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| namespace_name | texto | Nombre del espacio de nombres donde existe una función especificada. |
| function_name | texto | El nombre de la función. |
| argument_types | texto | Cadena que representa el tipo de argumento de entrada de una función. |
| privilege_type | texto | Tipo de permiso. El valor posible es EXECUTE. |
| identity_id | entero | ID de la identidad. Los valores posibles son ID de usuario, ID de rol o ID de grupo. |
| identity_name | texto | Nombre de la identidad. |
| identity_type | texto | El tipo de la identidad. Los valores posibles son usuario, rol, grupo o público. |
| admin_option | booleano | Valor que indica si el usuario puede conceder el permiso a otros usuarios y roles. Siempre es false (falso) para el tipo de identidad de rol y grupo. |

## Consulta de ejemplo

En el siguiente ejemplo, se muestra el resultado de SVV_FUNCTION_PRIVILEGES.

```
SELECT
 namespace_name,function_name,argument_types,privilege_type,identity_name,identity_type,admin_o
 FROM svv_function_privileges
WHERE identity_name IN ('role1', 'reguser');

 namespace_name | function_name |       argument_types       | privilege_type |
 identity_name | identity_type | admin_option
---------------+---------------+----------------------------+----------------
+----------------+---------------+-------------
    public      | test_func1    | integer                    |    EXECUTE     |
 role1     |     role      |  False
    public      | test_func2    | integer, character varying |    EXECUTE     |
 reguser    |     user      |  False
```

# SVV_GEOGRAPHY_COLUMNS

Utilice SVV_GEOGRAPHY_COLUMNS para ver la lista de columnas GEOGRAPHY en el almacenamiento de datos. Esta lista de columnas incluye columnas de recursos compartidos de datos.

SVV_GEOGRAPHY_COLUMNS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| f_table_catalog | varchar(128) | El nombre de la base de datos en la que se encuentra la tabla con la columna GEOGRAPHY. |
| f_table_schema | varchar(128) | El nombre del esquema en el que se encuentra la tabla con la columna GEOGRAPHY. |
| f_table_name | varchar(128) | El nombre de la tabla en la que se encuentra la columna GEOGRAPHY. |
| f_geography_column | varchar(128) | El nombre de la columna GEOGRAPHY. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| coord_dimension | entero | El número de dimensiones de los datos GEOGRAPHY. |
| srid | entero | El identificador del sistema de referencia espacial (SRID) de los datos GEOGRAPHY. |
| type | varchar(128) | El nombre del tipo de datos de geografía espacial. |

### Consulta de ejemplo

En el siguiente ejemplo se muestra el resultado de SVV_GEOGRAPHY_COLUMNS.

```
SELECT * FROM svv_geography_columns;

f_table_catalog  | f_table_schema  | f_table_name  | f_geography_column  |
 coord_dimension | srid |  type
-----------------+-----------------+---------------+---------------------
+-----------------+------+--------------
dev              | public          | spatial_test  | test_geography      | 2
    | 0    | GEOGRAPHY
```

## SVV_GEOMETRY_COLUMNS

Utilice SVV_GEOMETRY_COLUMNS para ver la lista de columnas GEOMETRY en el almacenamiento de datos. Esta lista de columnas incluye columnas de recursos compartidos de datos.

SVV_GEOMETRY_COLUMNS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| f_table_catalog | varchar(128) | El nombre de la base de datos en la que se encuentra la tabla con la columna GEOMETRY. |
| f_table_schema | varchar(128) | El nombre del esquema en el que se encuentra la tabla con la columna GEOMETRY. |
| f_table_name | varchar(128) | El nombre de la tabla en la que se encuentra la columna GEOMETRY. |
| f_geometry_column | varchar(128) | El nombre de la columna GEOMETRY. |
| coord_dimension | entero | El número de dimensiones de los datos GEOMETRY. |
| srid | entero | El identificador del sistema de referencia espacial (SRID) de la columna GEOMETRY. |
| type | varchar(128) | El nombre del tipo de geometría espacial. |

## Consulta de ejemplo

En el siguiente ejemplo se muestra el resultado de SVV_GEOMETRY_COLUMNS.

```
SELECT * FROM svv_geometry_columns;

f_table_catalog  | f_table_schema  | f_table_name  | f_geometry_column   |
 coord_dimension | srid |  type
-----------------+-----------------+---------------+---------------------
+-----------------+------+-------------
dev              | public          | accomodations | shape               | 2
    | 0    | GEOMETRY
```

```
dev              | public          | zipcode       | wkb_geometry        | 2
    | 0    | GEOMETRY
```

# SVV_IAM_PRIVILEGES

Utilice SVV_IAM_PRIVILEGES para ver los privilegios de IAM concedidos explícitamente a usuarios, roles y grupos.

SVV_IAM_PRIVILEGES es visible para los siguientes usuarios:

- Superusuarios
- Usuarios con el permiso ACCESS SYSTEM TABLE

Los demás usuarios solo pueden ver las entradas a las que tienen acceso.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| iam_arn | texto | Nombre del espacio de nombres. |
| commanc ype | texto | Tipos de privilegios. Los valores posibles son COPY, UNLOAD, CREATE MODEL o EXTERNAL FUNCTION. |
| identity_ id | entero | ID de identidad. Los valores posibles son ID de usuario, ID de rol o ID de grupo. |
| identity_ name | texto | Nombre de identidad. |
| identity_ type | texto | Tipo de identidad Los valores posibles son usuario, rol, grupo o público. |

## Consultas de ejemplo

En el siguiente ejemplo se muestran los resultados de SVV_IAM_PRIVILEGES.

```
SELECT * from SVV_IAM_PRIVILEGES ORDER BY IDENTITY_ID;
       iam_arn        | command_type | identity_id | identity_name | identity_type
----------------------+--------------+-------------+---------------+---------------
 default-aws-iam-role | COPY         |           0 | public        | public
 default-aws-iam-role | UNLOAD       |           0 | public        | public
 default-aws-iam-role | CREATE MODEL |           0 | public        | public
 default-aws-iam-role | EXFUNC       |           0 | public        | public
 default-aws-iam-role | COPY         |         106 | u1            | user
 default-aws-iam-role | UNLOAD       |         106 | u1            | user
 default-aws-iam-role | CREATE MODEL |         106 | u1            | user
 default-aws-iam-role | EXFUNC       |         106 | u1            | user
 default-aws-iam-role | COPY         |      118413 | r1            | role
 default-aws-iam-role | UNLOAD       |      118413 | r1            | role
 default-aws-iam-role | CREATE MODEL |      118413 | r1            | role
 default-aws-iam-role | EXFUNC       |      118413 | r1            | role
(12 rows)
```

# SVV_IDENTITY_PROVIDERS

La vista de SVV_IDENTITY_PROVIDERS devuelve el nombre y las propiedades adicionales de los proveedores de identidades. Para obtener más información sobre cómo crear un proveedor de identidades, consulte CREATE IDENTITY PROVIDER.

SVV_IDENTITY_PROVIDERS solo es visible para los superusuarios. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| uid | entero | El ID único del proveedor de identidades registrado. |
| nombre | texto | El nombre del proveedor de identidad. |
| type | texto | El tipo de proveedor de identidad. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| instanceid | texto | El diferenciador único entre instancias del mismo tipo. |
| namespc | texto | El prefijo de espacio de nombres del proveedor de identidades. |
| params | texto | El objeto JSON con parámetros para el proveedor de identidades. |
| enabled | bool | Indica si el proveedor de identidades está habilitado. |

Consultas de ejemplo

Para ver las propiedades del proveedor de identidades, ejecute una consulta como la siguiente después de crear los proveedores de identidades.

```
SELECT name, type, instanceid, namespc, params, enabled
FROM svv_identity_providers
ORDER BY 1;
```

La salida de muestra incluye descripciones de parámetros.

```
       name       | type  |              instanceid              | namespc |

                                      params

                     | enabled
------------------+-------+--------------------------------------+---------
+-----------------------------------------------------------------------------------------------------
+---------
 rs5517_azure_idp | azure | e40d4bb2-7670-44ae-bfb8-5db013221d73 | abc     |
 {"issuer":"https://login.microsoftonline.com/e40d4bb2-7670-44ae-bfb8-5db013221d73/
v2.0", "client_id":"871c010f-5e61-4fb1-83ac-98610a7e9110", "client_secret":,
 "audience":["https://analysis.windows.net/powerbi/connector/AmazonRedshift", "https://
analysis.windows.net/powerbi/connector/AWSRDS"]} | t
```

```
(1 row)
```

# SVV_INTEGRATION

SVV_INTEGRATION muestra detalles sobre la configuración de las integraciones.

SVV_INTEGRATION solo es visible para los superusuarios. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Para obtener información sobre las integraciones sin ETL, consulte Zero-ETL integrations.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| integration_id | character (128) | Identificador asociado con la integración. |
| target_database | character (128) | Base de datos de Amazon Redshift que recibe los datos de integración. |
| origen | character (128) | Datos de origen de la integración. Los tipos posibles son `MySQL`, `PostgreSQL` , y `S3_EVENT_NOTIFICATIONS` . |
| estado | character (128) | Estado de la integración. Entre los valores posibles se incluyen: `PendingDbConnectState` , `SchemaDiscoveryState` , `CdcRefreshState` y `ErrorState` . |
| current_lag | bigint | Tiempo de retraso actual (milisegundos) entre el origen y el destino de la integración. |
| last_replicated_checkpoint | character (128) | Último punto de control replicado. |
| total_tables_replicated | entero | Número total de tablas que se encuentran actualmente en el estado replicado. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| total_tables_failed | entero | Número de tareas que actualmente tienen el estado de error. |
| creation_time | marca de tiempo | El tiempo (UTC) del momento en que se creó la integración. Se define como el momento en que se crea la base de datos de destino a partir de la integración. |
| refresh_interval | entero | El intervalo de tiempo aproximado, en segundos, para actualizar los datos del origen sin ETL a la base de datos de destino. |
| source_database | character (128) | El nombre de la base de datos de origen. |
| is_history_mode | booleano | Un valor de TRUE indica que el modo de historial está activado. Con FALSE se indica que el modo de historial está desactivado. |

## Consultas de ejemplo

El siguiente comando SQL muestra las integraciones definidas actualmente.

```
select * from svv_integration;

           integration_id              | target_database | source |      state
   | current_lag |      last_replicated_checkpoint     | total_tables_replicated |
 total_tables_failed |       creation_time       |  refresh_interval  | source_database
 | is_history_mode
--------------------------------------+-----------------+--------+-----------------
+-------------+--------------------------------------+-------------------------
+---------------------+--------------------------+--------------------
+-----------------+-----------------
  99108e72-1cfd-414f-8cc0-0216acefac77 |     perfdb      |  MySQL | CdcRefreshState |
  56606106  | {"txn_seq":9834,"txn_id":126597515} |           152          |
 0         | 2023-09-19 21:05:27.520299|      720           + mysourceetl     | f
```

# SVV_INTEGRATION_TABLE_MAPPING

SVV_INTEGRATION_TABLE_MAPPING muestra la asignación de la base de datos, el esquema, la tabla, la columna y el tipo de datos del origen al destino cuando el valor identificador de esos campos sea diferente.

> **Note**
>
> Esta vista solo se rellena para los siguientes tipos de integraciones sin ETL:
>
> - Aplicaciones de terceros de AWS Glue a Amazon SageMaker Lakehouse
> - Amazon DynamoDB a Amazon SageMaker Lakehouse
>
> Para obtener más información, consulte Zero-ETL integrations en la Guía para desarrolladores de AWS Glue.

La transformación de los valores de los identificadores del origen al destino sigue estas reglas:

- Una letra mayúscula se convierte en minúscula.
- Un carácter que no sea una letra minúscula, un dígito o un guion bajo (_) se convierte en un carácter de subrayado (_).
- Si hay un conflicto con un valor de identificador existente, se añade un identificador único universal (UUID) al nuevo identificador.
- Si el valor del identificador de origen es una palabra clave de Amazon Redshift, el sufijo `_redshift` se añade al nuevo identificador.

Tras la transformación, un carácter debe ser una letra minúscula, un dígito o un guion bajo (_) y coincidir con el patrón de expresiones regex `[a-z0-9_]`. Los siguientes ejemplos muestran las reglas de conversión:

| Origen | Destino | Notas |
|---|---|---|
| foo | foo | Sin transformación |
| Barras | barra | |

| Origen | Destino | Notas |
|---|---|---|
| fooBar | foobar | |
| foo1 | foo1 | Sin transformación |
| foo_1 | foo_1 | Sin transformación |
| Bar @1 | bar_1 | |
| foo_bar@ | foo_bar_ | |
| caso | case_redshift | |

SVV_INTEGRATION_TABLE_MAPPING es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Para obtener información sobre las integraciones sin ETL, consulte Zero-ETL integrations en la Guía de administración de Amazon Redshift.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| integration_id | character (128) | Identificador asociado con la integración. |
| source_database | character (128) | El nombre de la base de datos en el origen. |
| target_database | character (128) | Base de datos de Amazon Redshift que recibe los datos de integración. |
| source_schema_name | character (128) | El nombre del esquema en el origen. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| target_schema_name | character (128) | Esquema de Amazon Redshift que recibe los datos de integración. |
| source_table_name | character (128) | El nombre de la tabla en el origen. |
| target_table_name | character (128) | Tabla de Amazon Redshift que recibe los datos de integración. |
| source_column_name | character (128) | El nombre de la columna en el origen. |
| target_column_name | character (128) | La columna de Amazon Redshift que recibe los datos de integración. |
| source_data_type | character (128) | El tipo de datos de la columna en el origen. |
| target_data_type | character (128) | El tipo de datos de la columna de Amazon Redshift que recibe los datos de integración. |

## Consultas de ejemplo

El siguiente comando SQL muestra la asignación de valores de metadatos del origen al destino.

```
select * from svv_integration_table_mapping;

           integration_id              | source_database | target_database |
 source_schema_name | target_schema_name | source_table_name | target_table_name |
--------------------------------------+-----------------+-----------------
+--------------------+-------------------+--------------------------------------+
  99108e72-1cfd-414f-8cc0-0216acefac77 |     mydatabase  |  mydatabase     |  myschema
         | myschema           | Mytable            | mytable           |


                                       | source_column_name | target_column_name |
 source_data_type | target_data_type |
```

```
                                        +--------------------+-------------------
+-------------------+------------------+
                                        | Mycolumnname       | mycolumnname       |
 Mydatatype       | mydatatype       |
```

# SVV_INTEGRATION_TABLE_STATE

SVV_INTEGRATION_TABLE_STATE muestra detalles sobre la información de integración en el nivel de tabla.

SVV_INTEGRATION_TABLE_STATE solo es visible para los superusuarios. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Para obtener más información, consulte Integraciones sin ETL.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| integration_id | character (128) | Identificador asociado con la integración. |
| target_database | character (128) | Se trata del nombre de la base de datos de Amazon Redshift. |
| schema_name | character (128) | Nombre del esquema de Amazon Redshift. |
| table_name | character (128) | El nombre de la tabla. |
| table_state | character (128) | Estado de la tabla. Los valores posibles son `Synced`, `Failed`, `Deleted`, `ResyncRequired` , `ResyncInitiated` y `DroppedSource` . El estado `DroppedSource` indica que el origen de una tabla del modo de historial se ha eliminado en el origen. |
| table_last_replicated_checkpoint | character (128) | Coordenadas del registro sincronizado actual. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| razón | character (256) | Motivo de la última transición de estado. Los motivos más comunes pueden ser tipos de datos no compatibles en las tablas, ya que las tablas no tienen claves principales. Para obtener más información sobre cómo solucionar problemas comunes, consulte Solución de problemas de integraciones sin ETL en Amazon Redshift. |
| last_updated_timestamp | TIMESTAMP sin zona horaria | Hora (UTC) en la que se actualizó la tabla por última vez. |
| table_rows | bigint | La cantidad total de filas en la tabla. |
| table_size | bigint | Tamaño de la tabla en megabytes (MB). |
| is_history_mode | booleano | Un valor de TRUE indica que el modo de historial está activado. Con FALSE se indica que el modo de historial está desactivado. |

## Consultas de ejemplo

El siguiente comando de SQL muestra las columnas del registro de las integraciones.

```
select * from svv_integration_table_state;

          integration_id              | target_database | schema_name |     table_name
   | table_state  |table_last_replicated_checkpoint | reason | last_updated_timestamp
   |table_rows  | table_size | is_history_mode
---------------------------------------+-----------------+-------------
+------------------+-------------+---------------------------------+--------
+---------------------------+------------+------------+-----------------
 4798e675-8f9f-4686-b05f-92c538e19629 |  sample_test2   |    sample   |
 SampleTestChannel |    Synced    |   {"txn_seq":3,"txn_id":3122}   |        |
 2023-05-12 12:40:30.656625 | 2          |   16       | f
```

# SVV_INTERLEAVED_COLUMNS

Utilice la vista SVV_INTERLEAVED_COLUMNS para identificar, de manera más fácil, si una tabla que utiliza claves de ordenación con intervalos debería volver a indizarse utilizando un comando VACUUM REINDEX. Para obtener más información acerca de cómo determinar con qué frecuencia ejecutar el comando VACUUM y cuándo ejecutar el comando VACUUM REINDEX, consulte Cómo minimizar los tiempos de limpieza.

Solo los superusuarios pueden ver SVV_INTERLEAVED_COLUMNS. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| tbl | entero | ID de la tabla. |
| col | entero | Índice con base cero para la columna. |
| interleaved_skew | numeric (19,2) | Relación que indica la cantidad de sesgos presentes en las columnas de clave de ordenación con intervalo de una tabla. Un valor de 1,00 indica que no hay sesgos, mientras que valores mayores indican más sesgos. Las tablas con un sesgo grande deberían volver a indizarse con el comando VACUUM REINDEX. |
| last_reindex | marca de tiempo | Hora en que se ejecutó por última vez el comando VACUUM REINDEX para la tabla especificada. Este valor es NULL si la tabla nunca se volvió a indizar o si se ha rotado la tabla subyacente de registros del sistema, STL_VACUUM, desde la última reindexación. |

## Consultas de ejemplo

Para identificar las tablas que podrían necesitar volver a indexarse, ejecute la siguiente consulta.

```
select tbl as tbl_id, stv_tbl_perm.name as table_name,
col, interleaved_skew, last_reindex
```

```
from svv_interleaved_columns, stv_tbl_perm
where svv_interleaved_columns.tbl = stv_tbl_perm.id
and interleaved_skew is not null;

 tbl_id | table_name | col | interleaved_skew | last_reindex
--------+------------+-----+------------------+---------------------
 100068 | lineorder  |   0 |             3.65 | 2015-04-22 22:05:45
 100068 | lineorder  |   1 |             2.65 | 2015-04-22 22:05:45
 100072 | customer   |   0 |             1.65 | 2015-04-22 22:05:45
 100072 | lineorder  |   1 |             1.00 | 2015-04-22 22:05:45
(4 rows)
```

# SVV_LANGUAGE_PRIVILEGES

Utilice SVV_LANGUAGE_PRIVILEGES para ver los permisos de lenguaje que tienen concedidos de forma explícita los usuarios, roles y grupos de la base de datos actual.

SVV_LANGUAGE_PRIVILEGES es visible para los siguientes usuarios:

- Superusuarios
- Usuarios con el permiso ACCESS SYSTEM TABLE

Los demás usuarios solo pueden ver las identidades a las que tienen acceso o que son de su propiedad.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| language_ name | texto | Nombre del lenguaje. |
| privilege_type | texto | Tipo de permiso. El valor posible es USAGE. |
| identity_id | entero | ID de la identidad. Los valores posibles son ID de usuario, ID de rol o ID de grupo. |
| identity_name | texto | Nombre de la identidad. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| identity_type | texto | El tipo de la identidad. Los valores posibles son usuario, rol, grupo o público. |
| admin_option | booleano | Valor que indica si el usuario puede conceder el permiso a otros usuarios y roles. Siempre es false (falso) para el tipo de identidad de rol y grupo. |

### Consulta de ejemplo

En el siguiente ejemplo, se muestra el resultado de SVV_LANGUAGE_PRIVILEGES.

```
SELECT language_name,privilege_type,identity_name,identity_type,admin_option FROM
 svv_language_privileges
WHERE identity_name IN ('role1', 'reguser');

 language_name | privilege_type | identity_name | identity_type | admin_option
---------------+----------------+---------------+---------------+--------------
   exfunc      |     USAGE      |    reguser    |     user      |    False
   exfunc      |     USAGE      |     role1     |     role      |    False
   plpythonu   |     USAGE      |    reguser    |     user      |    False
```

## SVV_MASKING_POLICY

Utilice SVV_MASKING_POLICY para ver todas las políticas de enmascaramiento creadas en el clúster.

Solo los superusuarios y los usuarios con el rol `sys:secadmin` pueden ver SVV_MASKING_POLICY. Los usuarios normales verán 0 filas.

### Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| policy_database | texto | Nombre de la base de datos en la que se ha creado la política de enmascaramiento. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| policy_name | texto | Nombre de la política de enmascaramiento. |
| input_columns | texto | Atributos proporcionados en la cláusula WITH de la instrucción CREATE POLICY. |
| policy_expression | texto | Expresión de enmascaramiento utilizada en la política. |
| policy_modified_by | texto | Nombre del usuario que modificó la política por última vez. |
| policy_modified_time | marca de tiempo | Marca temporal de cuando se creó o modificó por última vez la política. |

## SVV_ML_MODEL_INFO

Información acerca del estado actual del modelo de machine learning.

SVV_ML_MODEL_INFO es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

### Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| database_name | char(128) | La base de datos del modelo. |
| schema_name | char(128) | El esquema del modelo. |
| user_name | char(128) | El propietario del modelo. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| model_name | char(128) | El nombre del modelo. |
| life_cycle | char(20) | El estado del ciclo de vida del modelo. |
| is_refreshable | entero | Indica si el estado del modelo se puede actualizar en caso de que las tablas y las columnas originales en la consulta de formación sigan existiendo y de que el usuario aún cuente con los permisos necesarios para usarlas. Los valores posibles son 1 (actualizable) y 0 (no actualizable). |
| model_state | char(128) | El estado actual del modelo. |

### Consulta de ejemplo

La siguiente consulta muestra el estado actual de los modelos de machine learning.

```
SELECT schema_name, model_name, model_state
FROM svv_ml_model_info;

 schema_name |        model_name            |             model_state
-------------+------------------------------+-------------------------------------
 public      | customer_churn_auto_model    | Train Model On SageMaker In Progress
 public      | customer_churn_xgboost_model | Model is Ready
(2 row)
```

## SVV_ML_MODEL_PRIVILEGES

Utilice SVV_ML_MODEL_PRIVILEGES para ver los permisos de modelo de machine learning que tienen concedidos de forma explícita los usuarios, roles y grupos del clúster.

SVV_ML_MODEL_PRIVILEGES es visible para los siguientes usuarios:

- Superusuarios
- Usuarios con el permiso ACCESS SYSTEM TABLE

Los demás usuarios solo pueden ver las identidades a las que tienen acceso o que son de su propiedad.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| namespace_name | texto | Nombre del espacio de nombres donde existe un modelo de machine learning especificado. |
| model_name | texto | Nombre del modelo de machine learning. |
| model_version | entero | Número de versión del modelo. |
| privilege_type | texto | Tipo de permiso. El valor posible es EXECUTE. |
| identity_id | entero | ID de la identidad. Los valores posibles son ID de usuario, ID de rol o ID de grupo. |
| identity_name | texto | Nombre de la identidad. |
| identity_type | texto | El tipo de la identidad. Los valores posibles son usuario, rol, grupo o público. |
| admin_option | booleano | Valor que indica si el usuario puede conceder el permiso a otros usuarios y roles. Siempre es false (falso) para el tipo de identidad de rol y grupo. |

## Consulta de ejemplo

En el siguiente ejemplo, se muestra el resultado de SVV_ML_MODEL_PRIVILEGES.

```
SELECT
 namespace_name,model_name,model_version,privilege_type,identity_name,identity_type,admin_optio
 FROM svv_ml_model_privileges
WHERE model_name = 'test_model';

 namespace_name | model_name | model_version | privilege_type |  identity_name |
 identity_type | admin_option
---------------+------------+---------------+----------------+---------------
+---------------+--------------
      public    | test_model |       1       |    EXECUTE     |     reguser    |
 user      |    False
      public    | test_model |       1       |    EXECUTE     |     role1      |
 role      |    False
```

# SVV_MV_DEPENDENCY

En la tabla SVV_MV_DEPENDENCY, se muestran las dependencias de vistas materializadas con respecto a otras vistas materializadas de Amazon Redshift.

Para obtener más información acerca de las vistas materializadas, consulte Vistas materializadas en Amazon Redshift.

SVV_MV_DEPENDENCY es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| database_name | char(128) | La base de datos que contiene la vista materiali zada que se especificó. |
| schema_name | char(128) | Esquema de la vista materializada. |
| nombre | char(128) | El nombre de la vista materializada. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| dependent_database_name | char(128) | La base de datos de vista materializada de que la depende esta vista materializada. |
| dependent_schema_name | char(128) | El esquema del cual depende la vista materializada. |
| dependent_name | char(128) | El nombre del cual depende la vista materializada. |

### Consulta de ejemplo

La siguiente consulta devuelve una fila de salida que indica que la vista materializada `mv_over_foo` utiliza la vista materializada `mv_foo` en su definición como una dependencia.

```
CREATE SCHEMA test_ivm_setup;
CREATE TABLE test_ivm_setup.foo(a INT);
CREATE MATERIALIZED VIEW test_ivm_setup.mv_foo AS SELECT * FROM test_ivm_setup.foo;
CREATE MATERIALIZED VIEW test_ivm_setup.mv_over_foo AS SELECT * FROM
 test_ivm_setup.mv_foo;

SELECT * FROM svv_mv_dependency;

 database_name | schema_name          | name        | dependent_database_name |
 dependent_schema_name     | dependent_name
---------------+----------------------+-------------+-------------------------
+--------------------------+----------
 dev           | test_ivm_setup       | mv_over_foo |                     dev |
 test_ivm_setup            | mv_foo
```

## SVV_MV_INFO

La tabla SVV_MV_INFO contiene una fila para cada vista materializada, si los datos están o no obsoletos e información de estado.

Para obtener más información acerca de las vistas materializadas, consulte Vistas materializadas en Amazon Redshift.

Todos los usuarios pueden ver SVV_MV_INFO. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| database_name | char(128) | Base de datos que contiene la vista materializada. |
| schema_name | char(128) | El esquema de la base de datos. |
| user_name | char(128) | El usuario propietario de la vista materializada. |
| nombre | char(128) | El nombre de la vista materializada. |
| is_stale | char(1) | Una `t` indica que la vista materializada está obsoleta. Una vista materializada obsoleta es aquella en la que las tablas base se han actualiza do pero la vista materializada no. Esta informaci ón puede no ser precisa si no se ha ejecutado una actualización desde el último reinicio. |
| estado | entero | El estado de la vista materializada es el siguiente:<br>• 0: la vista materializada se vuelve a calcular por completo cada vez que se actualiza.<br>• 1: la vista materializada es progresiva.<br>• 101: la vista materializada no se puede actualizar debido a que se ha eliminado una columna. Esta restricción se aplica incluso si la columna no se utiliza en la vista materializada.<br>• 102: la vista materializada no se puede actualizar debido al cambio en el tipo de una columna. Esta restricción se aplica incluso si la columna no se utiliza en la vista materializada. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | • 103: la vista materializada no se puede actualizar debido a que se ha cambiado el nombre de una tabla.<br>• 104: la vista materializada no se puede actualizar debido a que se ha cambiado el nombre de una columna. Esta restricción se aplica incluso si la columna no se utiliza en la vista materializada.<br>• 105: la vista materializada no se puede actualizar debido a que se ha cambiado el nombre de un esquema. |
| autorewrite | char(1) | `t` indica que la vista materializada reúne los requisitos necesarios para la reescritura automática de las consultas. |
| autorefresh | char(1) | `t` indica que la vista materializada se puede actualizar de manera automática. |

## Consulta de ejemplo

Para ver el estado de todas las vistas materializadas, ejecute la siguiente consulta.

```
select * from svv_mv_info;
```

Esta consulta devuelve el siguiente ejemplo de salida.

```
database_name |       schema_name       | user_name |   name  |  is_stale | state |
 autorefresh | autorewrite
--------------+-------------------------+-----------+---------+-----------+-------
+-------------+----------------
 dev          | test_ivm_setup          | catch-22  | mv      |   f       |     1 |
       1 |           0
```

```
dev          | test_ivm_setup          | lotr      | old_mv  |   t       |     1 |
      0 |           1
```

# SVV_QUERY_INFLIGHT

Utilice la vista SVV_QUERY_INFLIGHT para determinar cuáles son las consultas que se están ejecutando actualmente en la base de datos. La vista combina STV_INFLIGHT con STL_QUERYTEXT. SVV_QUERY_INFLIGHT no muestra las consultas únicas del nodo principal. Para obtener más información, consulte Funciones específicas del nodo principal.

SVV_QUERY_INFLIGHT es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> Esta vista solo está disponible cuando se consultan clústeres aprovisionados.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| slice | entero | Sector donde se ejecuta la consulta. |
| consulta | entero | ID de la consulta. Se puede usar para combinar distintas vistas y tablas del sistema. |
| pid | entero | ID de proceso. Todas las consultas en una sesión se ejecutan en el mismo proceso, por lo que este valor permanece constante si ejecuta una serie de consultas en la misma sesión. Puede utilizar esta columna para combinar la tabla STL_ERROR. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| starttime | marca de tiempo | La hora en que comenzó la consulta. |
| suspended | entero | Indica si la consulta está suspendida: **0** = falso; **1** = verdadero. |
| texto | character(200) | Texto de la consulta, en incrementos de 200 caracteres. |
| sequence | entero | Número de secuencia para los segmentos de las instrucciones de la consulta. |

### Consultas de ejemplo

En el siguiente ejemplo de salida, se muestran dos consultas que están actualmente en ejecución, incluida la misma consulta SVV_QUERY_INFLIGHT y la consulta 428, que está dividida en tres filas en la tabla. (En este ejemplo de salida, las columnas starttime y statement están truncadas).

```
select slice, query, pid, starttime, suspended, trim(text) as statement, sequence
from svv_query_inflight
order by query, sequence;

slice|query| pid  |      starttime       |suspended| statement | sequence
-----+-----+------+----------------------+---------+-----------+---------
1012 | 428 | 1658 | 2012-04-10 13:53:... |       0 | select ...|    0
1012 | 428 | 1658 | 2012-04-10 13:53:... |       0 | enueid ...|    1
1012 | 428 | 1658 | 2012-04-10 13:53:... |       0 | atname,...|    2
1012 | 429 | 1608 | 2012-04-10 13:53:... |       0 | select ...|    0
(4 rows)
```

## SVV_QUERY_STATE

Utilice SVV_QUERY_STATE para ver información sobre el tiempo de ejecución de las consultas en ejecución actualmente.

La vista SVV_QUERY_STATE tiene un subconjunto de datos de la tabla STV_EXEC_STATE.

SVV_QUERY_STATE es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

> **Note**
>
> Esta vista solo está disponible cuando se consultan clústeres aprovisionados.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. Se puede usar para combinar distintas vistas y tablas del sistema. |
| seg | entero | Número del segmento de la consulta que se está ejecutando. Una consulta consta de distintos segmentos y cada segmento consta de uno o más pasos. Los segmentos de una consulta se ejecutan en paralelo. Cada segmento se ejecuta en un solo proceso. |
| paso | entero | Número del paso de la consulta que se está ejecutando. Un paso es la unidad mínima de tiempo de ejecución de una consulta. Cada paso representa una unidad discreta de trabajo, como el examen de una tabla, los resultados devueltos o el ordenamiento de datos. |
| maxtime | intervalo | Cantidad máxima de tiempo (en microsegundos) en que se ejecuta este paso. |
| avgtime | interval | Tiempo promedio (en microsegundos) en que se ejecuta este paso. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| rows | bigint | Número de filas producidas por el paso que se está ejecutando. |
| bytes | bigint | Número de bytes producidos por el paso que se está ejecutando. |
| cpu | bigint | Para uso interno. |
| memory | bigint | Para uso interno. |
| rate_row | double precision | Relación de filas por segundo desde que se inició la consulta, computadas al sumar las filas y dividirlas en la cantidad de segundos desde que se inició la consulta hasta el momento actual. |
| rate_byte | double precision | Relación de bytes por segundo desde que se inició la consulta, computadas al sumar los bytes y dividirlos en la cantidad de segundos desde que se inició la consulta hasta el momento actual. |
| label | character (25) | Etiqueta de la consulta: un nombre para el paso, como `scan` o `sort`. |
| is_diskbased | character (1) | Indica si este paso de la consulta se está ejecutando como una operación basada en disco: true (**`t`**) (verdadero) o false (**`f`**) (falso). Solo determinados pasos, como hash, ordenar y agregar pasos, pueden ir al disco. Muchos tipos de pasos se realizan siempre en la memoria. |
| workmem | bigint | Cantidad de memoria funcional (en bytes) asignada a este paso de la consulta. |
| num_parts | entero | Cantidad de particiones en que se divide una tabla hash durante un paso hash. Un número positivo en esta columna no significa que el paso hash se haya ejecutado como una operación basada en el disco. Controle el valor en la columna IS_DISKBASED para saber si el paso hash se basó en el disco. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| is_rrscan | character (1) | Si es true (**t**), indica que se utilizó un examen de rango restringido en el paso. El valor predeterminado es false (**f**). |
| is_delayed_scan | character (1) | Si es true (**t**), indica que en el paso se utilizó un examen diferido. El valor predeterminado es false (**f**). |

## Consultas de ejemplo

Determinar cuál es el tiempo de procesamiento de una consulta en cada paso

En la siguiente consulta, se muestra cuánto tiempo demoró en ejecutarse cada paso de la consulta con el ID de consulta 279 y cuántas filas de datos procesó Amazon Redshift:

```
select query, seg, step, maxtime, avgtime, rows, label
from svv_query_state
where query = 279
order by query, seg, step;
```

Esta consulta recupera la información de procesamiento acerca de la consulta 279, tal como se muestra en el siguiente ejemplo de salida:

```
query |   seg   | step | maxtime | avgtime |  rows   | label
------+---------+------+---------+---------+---------+------------------
  279 |       3 |    0 | 1658054 | 1645711 | 1405360 | scan
  279 |       3 |    1 | 1658072 | 1645809 |       0 | project
  279 |       3 |    2 | 1658074 | 1645812 | 1405434 | insert
  279 |       3 |    3 | 1658080 | 1645816 | 1405437 | distribute
  279 |       4 |    0 | 1677443 | 1666189 | 1268431 | scan
  279 |       4 |    1 | 1677446 | 1666192 | 1268434 | insert
  279 |       4 |    2 | 1677451 | 1666195 |       0 | aggr
(7 rows)
```

Determinar qué consultas activas se están ejecutando actualmente en el disco

En la siguiente consulta, se muestra si alguna de las consultas activas se está ejecutando actualmente en el disco:

```
select query, label, is_diskbased from svv_query_state
where is_diskbased = 't';
```

En este ejemplo de salida, se muestra si alguna de las consultas activas se está ejecutando actualmente en el disco:

```
 query | label        | is_diskbased
-------+--------------+--------------
1025   | hash tbl=142 |      t
(1 row)
```

# SVV_REDSHIFT_COLUMNS

Utilice SVV_REDSHIFT_COLUMNS para ver una lista de todas las columnas a las que tiene acceso un usuario. Este conjunto de columnas incluye las columnas del clúster y las columnas de los datashares proporcionados por los clústeres remotos.

SVV_REDSHIFT_COLUMNS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| database_name | varchar(128) | El nombre de la base de datos donde se encuentra la tabla que contiene las columnas. |
| schema_name | varchar(128) | Nombre del esquema para la tabla. |
| table_name | varchar(128) | El nombre de la tabla. |
| column_name | varchar(128) | El nombre de una columna. |
| ordinal_position | entero | La posición de la columna en la tabla. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| data_type | varchar (32) | El tipo de datos de la columna. |
| column_default | varchar(4000) | El valor predeterminado de la columna. |
| is_nullable | varchar(3) | Un valor que define si una columna puede contener valores nulos. Los valores posibles son `yes`, no y “ ” (una cadena vacía que no representa ninguna informaci ón). |
| encoding | varchar(128) | El tipo de codificación de la columna. |
| distkey | booleano | Un valor que es true si esta columna es la clave de distribución para la tabla y false en caso contrario. |

| Nombre de la columna | Tipo de datos: | Descripción |
| --- | --- | --- |
| sortkey | entero | Un valor que especifica el orden de la columna en la clave de ordenación.<br><br>Si la tabla utiliza una clave de ordenación compuesta , todas las columnas que forman parte de la clave de ordenación tienen un valor positivo que indica la posición de la columna en la clave de ordenación.<br><br>Si la tabla utiliza una clave de ordenación intercalada, cada una de las columnas que forman parte de la clave de ordenación tienen un valor que puede ser positivo o negativo. Aquí, el valor absoluto indica la posición de la columna en la clave de ordenación.<br><br>Si `sortkey` es 0, la columna no forma parte de la clave de ordenación. |
| column_acl | varchar(128) | Cadena que define los permisos del usuario o grupo de usuarios especificados para la columna. |
| remarks | varchar (256) | Observaciones. |

## Consulta de ejemplo

El siguiente ejemplo devuelve el resultado de SVV_REDSHIFT_COLUMNS.

```
SELECT *
FROM svv_redshift_columns
WHERE database_name = 'tickit_db'
    AND TABLE_NAME = 'tickit_sales_redshift'
ORDER BY COLUMN_NAME,
    TABLE_NAME,
    database_name
LIMIT 5;

database_name | schema_name |       table_name      | column_name | ordinal_position |
 data_type | column_default | is_nullable | encoding | distkey | sortkey | column_acl
 | remarks
--------------+-------------+-----------------------+-------------+------------------
+-----------+----------------+-------------+----------+---------+---------
+-------------+--------
   tickit_db  |   public    | tickit_sales_redshift |   buyerid   |        4         |
 integer  |                |      NO     |   az64   |  False  |    0    |             |
   tickit_db  |   public    | tickit_sales_redshift |  commission |        9         |
 numeric  |      (8,2)     |     YES     |   az64   |  False  |    0    |             |
   tickit_db  |   public    | tickit_sales_redshift |    dateid   |        6         |
 smallint |                |      NO     |   none   |  False  |    1    |             |
   tickit_db  |   public    | tickit_sales_redshift |   eventid   |        5         |
 integer  |                |      NO     |   az64   |  False  |    0    |       |
   tickit_db  |   public    | tickit_sales_redshift |   listid    |        2         |
 integer  |                |      NO     |   az64   |  True   |    0    |             |
```

# SVV_REDSHIFT_DATABASES

Utilice SVV_REDSHIFT_DATABASES para ver una lista de todas las bases de datos a las que tiene acceso un usuario. Esto incluye las bases de datos del clúster y las bases de datos creadas a partir de datashares proporcionados por clústeres remotos.

SVV_REDSHIFT_DATABASES es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| database_name | varchar(128) | El nombre de la base de datos. |
| database_owner | entero | El ID de usuario del propietario de la base de datos. |
| database_type | varchar (32) | El tipo de la base de datos. Los tipos posibles son bases de datos locales o compartidas. |
| database_acl | varchar(128) | Esta información es solo para uso interno. |
| database_options | varchar(128) | Las propiedades de la base de datos. |
| database_isolation_level | varchar(128) | El nivel de aislamiento de la base de datos. Los valores posibles son: `Snapshot Isolation` y `Serializable`. |

## Consulta de ejemplo

El siguiente ejemplo devuelve el resultado de SVV_REDSHIFT_DATABASES.

```
select database_name, database_owner, database_type, database_options,
 database_isolation_level
from  svv_redshift_databases;


database_name | database_owner | database_type | database_options |
 database_isolation_level
--------------+----------------+---------------+------------------+-----------------
```

```
dev        |  1             | local         | NULL              | Serializable
```

# SVV_REDSHIFT_FUNCTIONS

Utilice SVV_REDSHIFT_FUNCTIONS para ver una lista de todas las funciones a las que tiene acceso un usuario. Este conjunto de funciones incluye las funciones del clúster y las funciones de los datashares proporcionados por los clústeres remotos.

SVV_REDSHIFT_FUNCTIONS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| database_name | varchar(128) | El nombre de la base de datos donde se encuentra el clúster que tiene estas funciones. |
| schema_name | varchar(128) | El nombre del esquema que especifica una función dada. |
| function_name | varchar(128) | El nombre de una función especificada. |
| function_type | varchar(128) | El tipo de función. Los valores posibles son funciones regulares, funciones de agregación y procedimientos almacenados. |
| argument_type | varchar(512) | Cadena que representa el tipo de argumento de entrada de una función. |
| result_type | varchar(128) | El tipo de datos del valor de retorno de una función. |

### Consulta de ejemplo

El siguiente ejemplo devuelve el resultado de SVV_REDSHIFT_FUNCTIONS.

```
SELECT *
FROM svv_redshift_functions
WHERE database_name = 'tickit_db'
    AND SCHEMA_NAME = 'public'
ORDER BY function_name
LIMIT 5;

database_name | schema_name |      function_name    |  function_type   |
 argument_type  | result_type
--------------+-------------+-----------------------+------------------
+------------------+-------------
   tickit_db  |    public   |     shared_function   | REGULAR FUNCTION | integer,
 integer |   integer
```

## SVV_REDSHIFT_SCHEMA_QUOTA

Muestra la cuota y la utilización actual del disco para cada esquema de una base de datos.

SVV_REDSHIFT_SCHEMA_QUOTA es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Esta vista está disponible cuando se consultan clústeres aprovisionados o grupos de trabajo de Redshift sin servidor.

### Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| database_name | character (128) | La base de datos que contiene el esquema. |
| schema_name | character (128) | El nombre del esquema. |
| schema_owner | entero | El ID de usuario interno del propietario del esquema. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| cuota | entero | La cantidad de espacio en disco (en MB) que puede utilizar el esquema. |
| disk_usage | entero | El espacio en disco (en MB) que utiliza actualmente el esquema. |

### Consulta de ejemplo

En el siguiente ejemplo, se muestra la cuota y el uso de disco actual para el esquema denominado `sales_schema`.

```
SELECT TRIM(SCHEMA_NAME) "schema_name", QUOTA, disk_usage FROM
 svv_redshift_schema_quota
WHERE SCHEMA_NAME = 'sales_schema';


schema_name   | quota | disk_usage
--------------+-------+------------
sales_schema  | 2048  | 30
```

## SVV_REDSHIFT_SCHEMAS

Utilice SVV_REDSHIFT_SCHEMAS para ver una lista de todos los esquemas a los que tiene acceso un usuario. Este conjunto de esquemas incluye los esquemas del clúster y los esquemas de los datashares proporcionados por clústeres remotos.

SVV_REDSHIFT_SCHEMAS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| database_name | varchar(128) | El nombre de la base de datos donde se encuentra un esquema especificado. |
| schema_name | varchar(128) | El espacio de nombres o el nombre del esquema. |
| schema_owner | entero | El ID de usuario interno del propietario del esquema. |
| schema_type | varchar (16) | El tipo del esquema. Los valores posibles son esquemas compartidos y locales. |
| schema_acl | varchar(128) | Cadena que define los permisos del usuario o grupo de usuarios especificados para el esquema. |
| schema_option | varchar(128) | Las opciones del esquema. |

## Consulta de ejemplo

El siguiente ejemplo devuelve el resultado de SVV_REDSHIFT_SCHEMAS.

```
SELECT *
FROM svv_redshift_schemas
WHERE database_name = 'tickit_db'
ORDER BY database_name,
    SCHEMA_NAME;

database_name |    schema_name     | schema_owner | schema_type | schema_acl |
 schema_option
```

```
-------------+--------------------+-------------+------------+-----------
+---------------
   tickit_db  |       public       |       1     |    shared  |           |
```

# SVV_REDSHIFT_TABLES

Utilice SVV_REDSHIFT_TABLES para ver una lista de todas las tablas a las que tiene acceso un usuario. Este conjunto de tablas incluye las tablas del clúster y las tablas de los datashares proporcionados por los clústeres remotos.

SVV_REDSHIFT_TABLES es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| database_name | varchar(128) | El nombre de la base de datos donde existe una tabla especificada. |
| schema_name | varchar(128) | Nombre del esquema para la tabla. |
| table_name | varchar(128) | El nombre de la tabla. |
| table_type | varchar(128) | El tipo de tabla. Los valores posibles son vistas y tablas. |
| table_acl | varchar(128) | Cadena que define los permisos del usuario o grupo de usuarios especificados para la tabla. |
| remarks | varchar(128) | Observaciones. |
| table_owner | varchar(128) | El propietario de la tabla. |

### Consulta de ejemplo

El siguiente ejemplo devuelve el resultado de SVV_REDSHIFT_TABLES.

```
SELECT *
FROM svv_redshift_tables
WHERE database_name = 'tickit_db' AND TABLE_NAME LIKE 'tickit_%'
ORDER BY database_name,
TABLE_NAME;

database_name | schema_name |         table_name       | table_type | table_acl |
 remarks | table_owner
--------------+-------------+--------------------------+------------+-----------
+---------+-----------
   tickit_db  |    public   | tickit_category_redshift |    TABLE   |           |
   +
   tickit_db  |    public   |   tickit_date_redshift   |    TABLE   |           |
   +
   tickit_db  |    public   |   tickit_event_redshift  |    TABLE   |           |
   +
   tickit_db  |    public   |  tickit_listing_redshift |    TABLE   |           |
   +
   tickit_db  |    public   |   tickit_sales_redshift  |    TABLE   |           |
   +
   tickit_db  |    public   |   tickit_users_redshift  |    TABLE   |           |
   +
   tickit_db  |    public   |   tickit_venue_redshift  |    TABLE   |           |
```

Si el valor de table_acl es nulo, no se han concedido explícitamente privilegios de acceso a la tabla correspondiente.

## SVV_RELATION_PRIVILEGES

Utilice SVV_RELATION_PRIVILEGES para ver los permisos de relación (tablas y vistas) que tienen concedidos de forma explícita a los usuarios, roles y grupos de la base de datos actual.

SVV_RELATION_PRIVILEGES es visible para los siguientes usuarios:

- Superusuarios
- Usuarios con el permiso SYSLOG ACCESS UNRESTRICTED

Los demás usuarios solo pueden ver las identidades a las que tienen acceso o que son de su propiedad. Para obtener más información acerca de la visibilidad de datos, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| namespace_name | texto | Nombre del espacio de nombres donde existe una relación especificada. |
| relation_name | texto | Nombre de la relación. |
| privilege_type | texto | Tipo de permiso. Los valores posibles son INSERT, SELECT, UPDATE, DELETE, REFERENCES o DROP. |
| identity_id | entero | ID de la identidad. Los valores posibles son ID de usuario, ID de rol o ID de grupo. |
| identity_name | texto | Nombre de la identidad. |
| identity_type | texto | El tipo de la identidad. Los valores posibles son usuario, rol, grupo o público. |
| admin_option | booleano | Valor que indica si el usuario puede conceder el permiso a otros usuarios y roles. Siempre es false (falso) para el tipo de identidad de rol y grupo. |

## Consulta de ejemplo

En el siguiente ejemplo, se muestra el resultado de SVV_RELATION_PRIVILEGES.

```
SELECT
 namespace_name,relation_name,privilege_type,identity_name,identity_type,admin_option
 FROM svv_relation_privileges
WHERE relation_name = 'orders' AND privilege_type = 'SELECT';

 namespace_name | relation_name | privilege_type |  identity_name | identity_type |
 admin_option
```

```
---------------+--------------+---------------+---------------+--------------
+-------------
     public     |    orders     |     SELECT     |    reguser     |     user      |
 False
     public     |    orders     |     SELECT     |     role1      |     role      |
 False
```

# SVV_RLS_APPLIED_POLICY

Utilice SVV_RLS_APPLIED_POLICY para rastrear la aplicación de políticas de RLS en consultas que hacen referencia a relaciones protegidas por RLS.

SVV_RLS_APPLIED_POLICY es visible para los usuarios siguientes:

- Superusuarios
- Usuarios con el rol `sys:operator`
- Usuarios con el permiso ACCESS SYSTEM TABLE

Tenga en cuenta que sys:secadmin no tiene este permiso del sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| nombre de usuario | texto | El nombre del usuario que ejecutó la consulta. |
| consulta | entero | El Id. de la consulta. |
| xid | long | El contexto de la transacción. |
| pid | entero | El proceso principal que ejecuta la consulta. |
| recordtime | hora | La hora en la que se registró la consulta. |
| comando | char(1) | El comando para el que se aplicó la política de RLS. Los valores posibles son k para unknown, s para select, u para update, i para insert, y para utility y d para delete. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| datname | texto | El nombre de la base de datos de la relación a la que se adjunta la política de seguridad de la fila. |
| relschema | texto | El nombre del esquema de la relación a la que se adjunta la política de seguridad de la fila. |
| relname | texto | El nombre de la relación a la que se adjunta la política de seguridad de la fila. |
| polname | texto | El nombre de la política de seguridad de la fila que se adjunta a la relación. |
| poldefault | char(1) | La configuración predeterminada de la política de seguridad de la fila que se adjunta a la relación. Los valores posibles son `f` para false (falso) si se ha aplicado la política false predeterminada y `t` para true (verdadero) si se ha aplicado la política true predeterminada. |

## Consulta de ejemplo

En el ejemplo siguiente, se muestra el resultado de SVV_RLS_APPLIED_POLICY. Para consultar SVV_RLS_APPLIED_POLICY, debe contar con el permiso ACCESS SYSTEM TABLE.

```
-- Check what RLS policies were applied to the run query.
SELECT username, command, datname, relschema, relname, polname, poldefault
FROM svv_rls_applied_policy
WHERE datname = CURRENT_DATABASE() AND query = PG_LAST_QUERY_ID();

 username | command |  datname  | relschema |          relname         |      polname
  | poldefault
----------+---------+-----------+-----------+--------------------------
+-----------------+------------
   molly  |    s    | tickit_db |   public  | tickit_category_redshift |
 policy_concerts |
```

# SVV_RLS_ATTACHED_POLICY

Utilice SVV_RLS_ATTACHED_POLICY para ver una lista de todas las relaciones y los usuarios que tienen una o más políticas de seguridad de la fila adjuntas a la base de datos conectada actualmente.

Solo los usuarios con el rol sys:secadmin pueden consultar esta vista.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| relschema | texto | El nombre del esquema de la relación a la que se adjunta la política de seguridad de la fila. |
| relname | texto | El nombre de la relación a la que se adjunta la política de seguridad de la fila. |
| relkind | texto | El tipo de objeto, como tabla. |
| polname | texto | El nombre de la política de seguridad de la fila que se adjunta a la relación. |
| grantor | texto | El nombre del usuario que adjuntó la política. |
| grantee | texto | El nombre del usuario o rol al que se adjuntó la política. |
| granteekind | texto | El tipo de beneficiario. Los valores posibles son usuario o rol. |
| is_pol_on | booleano | El parámetro que indica si una política de seguridad de la fila está activada o desactivada en una tabla. Los valores posibles son true (verdadero) y false (falso). |
| is_rls_on | booleano | El parámetro que indica si una seguridad de la fila está activada o desactivada en una tabla. Los valores posibles son true (verdader o) y false (falso). |
| rls_conju nction_type | character (3) | Es el parámetro que indica si la relación combina las políticas de RLS con `and` o `or`. |

### Consulta de ejemplo

En el ejemplo siguiente, se muestra el resultado de SVV_RLS_ATTACHED_POLICY.

```
--Inspect the policy in SVV_RLS_ATTACHED_POLICY
SELECT * FROM svv_rls_attached_policy;

 relschema |        relname           | relkind |     polname     | grantor | grantee
 | granteekind | is_pol_on | is_rls_on | rls_conjuntion_type
-----------+--------------------------+---------+-----------------+---------+----------
+-------------+-----------+-----------+---------------------
 public    | tickit_category_redshift |  table  | policy_concerts |   bob   |  analyst
 |    role     |    True   |    True   |      and
 public    | tickit_category_redshift |  table  | policy_concerts |   bob   |  dbadmin
 |    role     |    True   |    True   |      and
```

## SVV_RLS_POLICY

Utilice SVV_RLS_POLICY para ver una lista de todas las políticas de seguridad de la fila creadas en el clúster de Amazon Redshift.

SVV_RLS_POLICY es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

### Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| poldb | texto | El nombre de la base de datos en la que se crea la política de seguridad de la fila. |
| polname | texto | El nombre de la política de seguridad de la fila. |
| polalias | texto | El alias de la tabla que se utiliza en la definición de la política. |
| polatts | texto | Los atributos proporcionados a la definición de la política. |
| polqual | texto | La condición de política proporcionada en la cláusula USING de la instrucción CREATE POLICY. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| polenabled | booleano | Si la política está activada en todos los casos. |
| polmodifiedby | texto | El nombre del último usuario que creó o modificó la política. |
| polmodifiedtime | marca de tiempo | La marca temporal del momento en que se creó o modificó la política por última vez. |

### Consulta de ejemplo

En el ejemplo siguiente, se muestra el resultado de SVV_RLS_POLICY.

```
-- Create some policies.
CREATE RLS POLICY pol1 WITH (a int) AS t USING ( t.a IS NOT NULL );
CREATE RLS POLICY pol2 WITH (c varchar(10)) AS t USING ( c LIKE '%public%');

-- Inspect the policy in SVV_RLS_POLICY
SELECT * FROM svv_rls_policy;

 poldb | polname | polalias |                     polatts                      |
         polqual                | polenabled | polmodifiedby |   polmodifiedtime
-------+---------+----------+--------------------------------------------------
+--------------------------------------+------------+---------------
+---------------------
 my_db | pol1    | t        | [{"colname":"a","type":"integer"}]               |
 "t"."a" IS NOT NULL                   | t          | policy_admin  | 2022-02-11
 14:40:49
 my_db | pol2    | t        | [{"colname":"c","type":"character varying(10)"}] |
 "t"."c" LIKE CAST('%public%' AS TEXT) | t          | policy_admin  | 2022-02-11
 14:41:28
```

## SVV_RLS_RELATION

Utilice SVV_RLS_RELATION para ver una lista de todas las relaciones que están protegidas por RLS.

SVV_RLS_RELATION es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| datname | texto | El nombre de la base de datos que contiene la relación. |
| relschema | texto | El nombre del esquema que contiene la relación. |
| relname | texto | Nombre de la relación. |
| relkind | texto | El tipo de relación, como tablas o vistas. |
| is_rls_on | booleano | El parámetro que indica si la relación está protegida por RLS. |
| is_rls_datashare_on | booleano | El parámetro que indica si la relación está protegida por RLS a través de recursos compartidos de datos. |
| rls_conjunction_type | character(3) | Es el parámetro que indica si la relación combina las políticas de RLS con `and` o `or`. |
| rls_datashare_conjunction_type | character(3) | Es el parámetro que indica si la relación combina las políticas de RLS con `and` o `or` en los recursos compartidos de datos. |

## Consulta de ejemplo

En el siguiente ejemplo, se muestra el resultado de SVV_RLS_RELATION.

```
ALTER TABLE tickit_category_redshift ROW LEVEL SECURITY ON FOR DATASHARES;


--Inspect RLS state on the relations using SVV_RLS_RELATION.
SELECT datname, relschema, relname, relkind, is_rls_on, is_rls_datashare_on FROM
 svv_rls_relation ORDER BY relname;
```

```
  datname  | relschema |        relname           | relkind | is_rls_on |
 is_rls_datashare_on | rls_conjunction_type | rls_datashare_conjunction_type
-----------+-----------+--------------------------+---------+-----------
+--------------------+----------------------+--------------------------------
 tickit_db |   public  | tickit_category_redshift |  table  |      t    |           t
       |          and         |              and
(1 row)
```

# SVV_ROLE_GRANTS

Utilice SVV_ROLE_GRANTS para ver una lista de roles que tienen concedidos roles de forma explícita en el clúster.

SVV_ROLE_GRANTS es visible para los usuarios siguientes:

- Superusuarios
- Usuarios con el permiso ACCESS SYSTEM TABLE

Los demás usuarios solo pueden ver las identidades a las que tienen acceso o que son de su propiedad.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| role_id | entero | ID del rol. |
| role_name | texto | Nombre del rol |
| granted_role_id | entero | ID del rol concedido. |
| granted_role_name | texto | Nombre del rol concedido. |

## Consulta de ejemplo

El siguiente ejemplo devuelve el resultado de SVV_ROLE_GRANTS.

```
GRANT ROLE role1 TO ROLE role2;
GRANT ROLE role2 TO ROLE role3;
```

```
SELECT role_name, granted_role_name FROM svv_role_grants;

 role_name |  granted_role_name
-----------+--------------------
   role2   |      role1
   role3   |      role2
(2 rows)
```

# SVV_ROLES

Utilice SVV_ROLES para ver la información de los roles.

Todos los usuarios pueden ver la tabla.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| role_id | entero | ID del rol. |
| role_name | texto | Nombre del rol |
| role_owner | texto | Nombre del propietario del rol. |
| external_id | texto | Identificador único del rol en el proveedor de identidades de terceros. |

## Consulta de ejemplo

El siguiente ejemplo devuelve el resultado de SVV_ROLES.

```
SELECT role_name,role_owner FROM svv_roles WHERE role_name IN ('role1', 'role2');

 role_name | role_owner
-----------+------------
   role1   | superuser
   role2   | superuser
```

# SVV_SCHEMA_PRIVILEGES

Utilice SVV_SCHEMA_PRIVILEGES para ver los permisos de esquema que tienen concedidos de forma explícita los usuarios, roles y grupos de la base de datos actual.

SVV_SCHEMA_PRIVILEGES es visible para los siguientes usuarios:

- Superusuarios
- Usuarios con el permiso ACCESS SYSTEM TABLE

Los demás usuarios solo pueden ver las identidades a las que tienen acceso o que son de su propiedad.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| namespace_name | texto | Nombre del espacio de nombres donde existe un esquema especific ado. |
| privilege_type | texto | Tipo de permiso. Para los permisos con privilege_scope de esquema, los valores posibles son CREATE, USAGE y ALTER. Para valores de privilege_scope distintos de esquema, los valores posibles incluyen cualquier tipo de permiso disponible en el ámbito del permiso. |
| identity_id | entero | ID de la identidad. Los valores posibles son ID de usuario, ID de rol o ID de grupo. |
| identity_name | texto | Nombre de la identidad. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| identity_type | texto | El tipo de la identidad. Los valores posibles son usuario, rol, grupo o público. |
| admin_option | booleano | Valor que indica si el usuario puede conceder el permiso a otros usuarios y roles. Siempre es false (falso) para el tipo de identidad de rol y grupo. |
| privilege_scope | texto | El alcance del permiso especific ado en privilege_type. Los valores posibles son los siguientes:<br>• SCHEMA<br>• TABLES<br>• FUNCTIONS<br><br>Para obtener información sobre los permisos acotados, vaya a Permisos acotados. |

## Consulta de ejemplo

En el siguiente ejemplo, se muestra el resultado de SVV_SCHEMA_PRIVILEGES.

```
SELECT namespace_name,privilege_type,identity_name,identity_type,admin_option FROM
 svv_schema_privileges
WHERE namespace_name = 'test_schema1';

 namespace_name | privilege_type |  identity_name | identity_type | admin_option
----------------+----------------+----------------+---------------+--------------
 test_schema1   |    USAGE       |     reguser    |     user      |   False
 test_schema1   |    USAGE       |     role1      |     role      |   False
```

# SVV_SCHEMA_QUOTA_STATE

Muestra la cuota y el uso actual del disco para cada esquema.

Los usuarios normales pueden ver información de esquemas para los que tienen permiso de USO. Los superusuarios pueden ver información de todos los esquemas de la base de datos actual.

SVV_SCHEMA_QUOTA_STATE es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> Esta vista solo está disponible cuando se consultan clústeres aprovisionados.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| schema_id | entero | El espacio de nombres o el ID de esquema. |
| schema_name | carácter (128) | El espacio de nombres o el nombre del esquema. |
| schema_owner | entero | El ID de usuario interno del propietario del esquema. |
| cuota | entero | La cantidad de espacio en disco (en MB) que puede utilizar el esquema. |
| disk_usage | entero | El espacio en disco (en MB) que utiliza actualmente el esquema. |
| disk_usage_pct | double precision | El porcentaje de espacio en disco que utiliza actualmente el esquema fuera de la cuota configurada. |

### Consulta de ejemplo

En el siguiente ejemplo, se muestra la cuota y el uso de disco actual para el esquema.

```
SELECT TRIM(SCHEMA_NAME) "schema_name", QUOTA, disk_usage, disk_usage_pct FROM
 svv_schema_quota_state
WHERE SCHEMA_NAME = 'sales_schema';
schema_name   | quota | disk_usage | disk_usage_pct
--------------+-------+------------+---------------
sales_schema  | 2048  | 30         | 1.46
(1 row)
```

## SVV_SYSTEM_PRIVILEGES

SVV_SYSTEM_PRIVILEGES es visible para los siguientes usuarios:

- Superusuarios
- Usuarios con el permiso ACCESS SYSTEM TABLE

Los demás usuarios solo pueden ver las identidades a las que tienen acceso o que son de su propiedad.

### Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| system_privilege | texto | Nombre del permiso del sistema. |
| identity_id | entero | ID de la identidad. Los valores posibles son ID de usuario o ID de rol. |
| identity_name | texto | Nombre de la identidad. |
| identity_type | texto | El tipo de la identidad. Los valores posibles son usuario o rol. |

### Consulta de ejemplo

En el siguiente ejemplo, se muestra el resultado de los parámetros especificados.

```
SELECT system_privilege,identity_name,identity_type FROM svv_system_privileges
WHERE system_privilege = 'ALTER TABLE' AND identity_name = 'sys:superuser';

 system_privilege | identity_name | identity_type
------------------+---------------+---------------
   ALTER TABLE    | sys:superuser |     role
```

## SVV_TABLE_INFO

Muestra información de resumen acerca de las tablas en la base de datos. La vista filtra las tablas de sistema y muestra solamente las tablas definidas por el usuario.

Puede utilizar la vista SVV_TABLE_INFO para diagnosticar y dar respuesta a problemas de diseño de las tablas que puedan influir en el rendimiento de las consultas. Eso incluye problemas con la codificación de compresión, las claves de distribución, el estilo de ordenación, el sesgo de la distribución de los datos, el tamaño de las tablas y las estadísticas. La vista SVV_TABLE_INFO no devuelve información para las tablas vacías.

En la vista SVV_TABLE_INFO, se resume la información de las tablas de sistema STV_BLOCKLIST, STV_NODE_STORAGE_CAPACITY, STV_TBL_PERM y STV_SLICES, así como de las tablas de catálogo PG_DATABASE, PG_ATTRIBUTE, PG_CLASS, PG_NAMESPACE y PG_TYPE.

Solo los superusuarios pueden ver SVV_TABLE_INFO. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema. Para permitir que un usuario pueda consultar la vista, conceda el permiso SELECT en SVV_TABLE_INFO al usuario.

### Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| database | texto | Nombre de la base de datos. |
| schema | texto | Nombre del esquema. |
| table_id | oid | ID de la tabla. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| table | texto | Nombre de la tabla. |
| encoded | texto | Valor que indica si la columna tiene definida alguna codificación de compresión. |
| diststyle | texto | Estilo de distribución o columna de clave de distribución, siempre que se haya definido una clave de distribución. Entre los valores posibles se incluyen EVEN, KEY(*column*), ALL, AUTO(ALL) , AUTO(EVEN) y AUTO(KEY( *column*)). |
| sortkey1 | texto | Primera columna en la clave de ordenación, siempre que se haya definido una clave de ordenación. Entre los valores posibles se incluyen: *column*, AUTO(SORTKEY) y AUTO(SORTKEY( *column*)). |
| max_varchar | entero | Tamaño de la columna más grande que utiliza un tipo de datos VARCHAR. |
| sortkey1_enc | character (32) | Codificación de compresión de la primera columna en la clave de ordenación, siempre que se haya definido una clave de ordenación. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| sortkey_num | entero | Cantidad de columnas definidas como claves de ordenación. |
| size | bigint | Tamaño de la tabla, en bloques de datos de 1 MB. |
| pct_used | numeric (10,4) | Porcentaje de espacio disponible que usa la tabla. |
| empty | bigint | Para uso interno. Esta columna ya no se utiliza y se eliminará en una futura versión. |
| unsorted | numeric (5,2) | Porcentaje de filas desordenadas en la tabla. |
| stats_off | numeric (5,2) | Número que indica cuán obsoletas son las estadísticas de la tabla; 0 es actual y 100 es desactualizada. |
| tbl_rows | numeric (38,0) | Cantidad total de filas en la tabla. Este valor incluye las filas marcadas para ser eliminadas, pero que aún no se han limpiado. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| skew_sortkey1 | numeric (19,2) | Proporción del tamaño de la columna más grande que no sea una clave de ordenación con el tamaño de la primera columna de la clave de ordenación, siempre que se haya definido una clave de ordenación. Utilice este valor para evaluar la eficacia de la clave de ordenación. |
| skew_rows | numeric (19,2) | Proporción de la cantidad de filas en el sector con la mayor cantidad de filas con la cantidad de filas en el sector con la menor cantidad de filas. |
| estimated_visible_rows | numeric (38,0) | Las filas estimadas en la tabla. Este valor no incluye las filas marcadas para ser eliminadas. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| risk_event | texto | Información de riesgo sobre una tabla. El campo está separado en partes:<br><br>`risk_type \|xid\|timestamp`<br><br>• `risk_type` , donde 1 indica que se ejecutó `COPY command with the EXPLICIT_IDS option`. Amazon Redshift ya no verifica la unicidad de las columnas IDENTITY en la tabla. Para obtener más información, consulte EXPLICIT_IDS.<br>• El ID de la transacción, `xid`, que introduce el riesgo.<br>• El `timestamp` cuando se ejecutó el comando COPY.<br><br>El siguiente ejemplo muestra el valor en el campo.<br><br>`1\|1107\|2019-06-22 07:16:11.292952` |
| vacuum_sort_benefit | numeric(12,2) | El porcentaje máximo de mejora estimado del rendimiento de la consulta del escaneo cuando ejecuta una ordenación vacuum. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| create_time | TIMESTAMP sin zona horaria | Marca de tiempo UNIX de cuando se creó la tabla. |

### Consultas de ejemplo

En el siguiente ejemplo, se muestra la codificación, el estilo de distribución, la ordenación y los sesgos de datos para todas las tablas definidas por el usuario en la base de datos. Aquí, "table" se debe indicar entre comillas dobles porque es una palabra reservada.

```
select "table", encoded, diststyle, sortkey1, skew_sortkey1, skew_rows
from svv_table_info
order by 1;

table          | encoded | diststyle       | sortkey1     | skew_sortkey1 | skew_rows
---------------+---------+-----------------+--------------+---------------+----------
category       | N       | EVEN            |              |               |
date           | N       | ALL             | dateid       |          1.00 |
event          | Y       | KEY(eventid)    | dateid       |          1.00 |      1.02
listing        | Y       | KEY(listid)     | dateid       |          1.00 |      1.01
sales          | Y       | KEY(listid)     | dateid       |          1.00 |      1.02
users          | Y       | KEY(userid)     | userid       |          1.00 |      1.01
venue          | N       | ALL             | venueid      |          1.00 |
(7 rows)
```

## SVV_TABLES

Utilice SVV_TABLES para ver las tablas en catálogos locales y externos.

SVV_TABLES es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

### Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| table_catalog | texto | Nombre del catálogo donde se encuentra la tabla. |
| table_schema | texto | Nombre del esquema para la tabla. |
| table_name | texto | El nombre de la tabla. |
| table_type | texto | El tipo de tabla. Los valores posibles son vistas, tablas externas y tablas base. |
| remarks | texto | Observaciones. |

## SVV_TRANSACTIONS

Registra información acerca de transacciones que actualmente tienen bloqueos en las tablas de la base de datos. Utilice la vista SVV_TRANSACTIONS para identificar las transacciones abiertas y bloquear los problemas de contención. Para obtener más información acerca de bloqueos, consulte Administración de operaciones de escritura simultáneas y LOCK.

SVV_TRANSACTIONS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

### Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| txn_owner | texto | El nombre del propietario de la transacción. |
| txn_db | texto | El nombre de la base de datos asociada a la transacción. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| xid | bigint | ID de transacción. |
| pid | entero | ID del proceso asociado al bloqueo. |
| txn_start | marca de tiempo | Hora de inicio de la transacción. |
| lock_mode | texto | Nombre del modo de bloqueo sostenido o solicitado por este proceso. Si `lock_mode` es `ExclusiveLock` y `granted` es true (`t`), entonces el ID de transacción es una transacción abierta. |
| lockable_object_type | texto | Tipo de objeto que solicita o sostiene el bloqueo, ya sea `relation` si se trata de una tabla o `transactionid` si se trata de una transacción. |
| relation | entero | ID de la tabla para la tabla (relación) que adquiere el bloqueo. Este valor es NULL si `lockable_object_type` es `transactionid` . |
| granted | booleano | Valor que indica si el bloqueo se ha concedido (`t`) o si está pendiente (`f`). |

## Consultas de ejemplo

Mediante el siguiente comando, se muestran todas las transacciones activas y los bloqueos solicitados por cada transacción.

```
select * from svv_transactions;

 txn_
 lockable_
 owner | txn_db |  xid   |  pid  |         txn_start          |      lock_mode      |
 object_type    | relation | granted
-------+--------+--------+-------+----------------------------+---------------------
+---------------+----------+---------
 root  | dev    | 438484 | 22223 | 2016-03-02 18:42:18.862254 | AccessShareLock     |
 relation       |   100068 | t
 root  | dev    | 438484 | 22223 | 2016-03-02 18:42:18.862254 | ExclusiveLock       |
 transactionid  |          | t
 root  | tickit | 438490 | 22277 | 2016-03-02 18:42:48.084037 | AccessShareLock     |
 relation       |    50860 | t
 root  | tickit | 438490 | 22277 | 2016-03-02 18:42:48.084037 | AccessShareLock     |
 relation       |    52310 | t
 root  | tickit | 438490 | 22277 | 2016-03-02 18:42:48.084037 | ExclusiveLock       |
 transactionid  |          | t
 root  | dev    | 438505 | 22378 | 2016-03-02 18:43:27.611292 | AccessExclusiveLock |
 relation       |   100068 | f
 root  | dev    | 438505 | 22378 | 2016-03-02 18:43:27.611292 | RowExclusiveLock    |
 relation       |    16688 | t
 root  | dev    | 438505 | 22378 | 2016-03-02 18:43:27.611292 | AccessShareLock     |
 relation       |   100064 | t
 root  | dev    | 438505 | 22378 | 2016-03-02 18:43:27.611292 | AccessExclusiveLock |
 relation       |   100166 | t
 root  | dev    | 438505 | 22378 | 2016-03-02 18:43:27.611292 | AccessExclusiveLock |
 relation       |   100171 | t
 root  | dev    | 438505 | 22378 | 2016-03-02 18:43:27.611292 | AccessExclusiveLock |
 relation       |   100190 | t
 root  | dev    | 438505 | 22378 | 2016-03-02 18:43:27.611292 | ExclusiveLock       |
 transactionid  |          | t
(12 rows)

(12 rows)
```

# SVV_USER_GRANTS

Utilice SVV_USER_GRANTS para ver la lista de usuarios que tienen concedidos roles de forma explícita en el clúster.

SVV_USER_GRANTS es visible para los siguientes usuarios:

- Superusuarios
- Usuarios con el permiso ACCESS SYSTEM TABLE

Los demás usuarios solo pueden ver los roles que se les conceden de forma explícita.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | ID del usuario. |
| user_name | texto | El nombre del usuario. |
| role_id | entero | ID de rol del rol concedido. |
| role_name | texto | Nombre de rol del rol concedido. |
| admin_option | booleano | Valor que indica si el usuario puede conceder el rol a otros usuarios y roles. |

## Consultas de ejemplo

Las siguientes consultas conceden roles a los usuarios y muestran la lista de usuarios que tienen concedidos roles de forma explícita.

```
GRANT ROLE role1 TO reguser;
GRANT ROLE role2 TO reguser;
GRANT ROLE role1 TO superuser;
GRANT ROLE role2 TO superuser;

SELECT user_name,role_name,admin_option FROM svv_user_grants;

 user_name | role_name | admin_option
-----------+-----------+--------------
 superuser |  role1    | False
 reguser   |  role1    | False
 superuser |  role2    | False
  reguser  |  role2    | False
```

# SVV_USER_INFO

Puede recuperar datos sobre los usuarios de la base de datos de Amazon Redshift con la vista SVV_USER_INFO.

SVV_USER_INFO es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_name | texto | Nombre de usuario de la función. |
| user_id | entero | ID del usuario. |
| createdb | booleano | Valor que indica si el usuario tiene permisos para crear bases de datos. |
| superusuario | booleano | Valor que indica si el usuario es un superusuario. |
| catalog_update | booleano | Valor que indica si el usuario puede actualizar los catálogos de sistema. |
| connection_limit | texto | Número de conexiones que el usuario puede abrir. |
| syslog_access | texto | Valor que indica si el usuario tiene acceso a los registros del sistema. Los dos valores posibles son `RESTRICTED` y `UNRESTRICTED`. `RESTRICTED` significa que los usuarios que no son superusuarios pueden ver sus propios registros. `UNRESTRICTED` significa que los usuarios que no son superusuarios pueden ver todos los registros de las vistas y tablas del sistema en las que tienen privilegios `SELECT`. |
| last_ddl_timestamp | marca de tiempo | Marca temporal de la instrucción de creación del lenguaje de definición de datos (DDL) ejecutada por el usuario. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| session_timeout | entero | El tiempo máximo en segundos durante el cual una sesión permanece inactiva o en reposo antes de que se cierre. El 0 indica que no se establece tiempo de espera. Para obtener información acerca de la configuración de tiempo de espera en reposo o inactivo del clúster, consulte Cuotas y límites de Amazon Redshift en la Guía de administración de Amazon Redshift. |
| external_user_id | texto | Identificador único del usuario en el proveedor de identidades de terceros. |

### Consultas de ejemplo

El siguiente comando recupera la información del usuario de SVV_USER_INFO.

```
SELECT * FROM SVV_USER_INFO;
```

## SVV_VACUUM_PROGRESS

Esta vista devuelve una estimación de cuánto tiempo demorará en completarse una operación de limpieza que actualmente está en progreso.

Solo los superusuarios pueden ver SVV_VACUUM_PROGRESS. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_VACUUM_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Para obtener más información sobre SVV_VACUUM_SUMMARY, consulte SVV_VACUUM_SUMMARY.

Para obtener más información sobre SVL_VACUUM_PERCENTAGE, consulte SVL_VACUUM_PERCENTAGE.

> **Note**
>
> Esta vista solo está disponible cuando se consultan clústeres aprovisionados.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| table_name | texto | Nombre de la tabla que actualmente se está limpiando o de la tabla que se limpió por última vez si es que no hay ninguna operación en progreso. |
| status | texto | Descripción de la actividad actual que se realiza como parte de la operación de limpieza:<br>• Initialize<br>• Ordenar<br>• Merge<br>• Eliminar<br>• Seleccionar<br>• Con error<br>• Completado<br>• Omitido<br>• Creación de un orden INTERLEAVED SORTKEY |
| time_remaining_estimate | texto | Tiempo restante estimado para que se complete la actual operación de limpieza, en minutos y segundos: por ejemplo, **`5m 10s`**. No se devuelve un tiempo estimado hasta que la operación de limpieza completa su primera operación de orden. Si no hay una operación de limpieza en curso, se muestra la última limpieza realizada con el valor **`Completed`** en la columna STATUS y la columna TIME_REMAINING_ESTIMATE vacía. Por lo general, la estimación se vuelve más precisa a medida que la operación de limpieza progresa. |

### Consultas de ejemplo

Las siguientes consultas, ejecutadas con unos pocos minutos de diferencia, muestran que se está limpiando una tabla grande llamada SALESNEW.

```
select * from svv_vacuum_progress;

table_name    |           status             | time_remaining_estimate
--------------+------------------------------+-------------------------
salesnew      |  Vacuum: initialize salesnew  |
(1 row)
...
select * from svv_vacuum_progress;

table_name   |         status         | time_remaining_estimate
-------------+------------------------+-------------------------
salesnew     |  Vacuum salesnew sort  | 33m 21s
(1 row)
```

La siguiente consulta muestra que actualmente no hay ninguna operación de limpieza en progreso. La última tabla que se limpió fue la tabla SALES.

```
select * from svv_vacuum_progress;

table_name   |  status  | time_remaining_estimate
-------------+----------+-------------------------
  sales      | Complete |
(1 row)
```

## SVV_VACUUM_SUMMARY

La vista SVV_VACUUM_SUMMARY combina las tablas STL_VACUUM, STL_QUERY y STV_TBL_PERM para resumir información relacionada con las operaciones de limpieza registradas por el sistema. La vista devuelve una fila por tabla por transacción de limpieza. La vista registra el tiempo transcurrido de la operación, la cantidad de particiones de orden creadas, la cantidad de incrementos de fusión requeridos y los deltas en conteos de fila y bloqueos antes y después de que se realiza la operación.

Solo los superusuarios pueden ver SVV_VACUUM_SUMMARY. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_VACUUM_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Para obtener más información sobre SVV_VACUUM_PROGRESS, consulte SVV_VACUUM_PROGRESS.

Para obtener más información sobre SVL_VACUUM_PERCENTAGE, consulte SVL_VACUUM_PERCENTAGE.

> **Note**
>
> Esta vista solo está disponible cuando se consultan clústeres aprovisionados.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| table_name | texto | Nombre de la tabla limpiada. |
| xid | bigint | ID de transacción de la operación VACUUM. |
| sort_partitions | bigint | Cantidad de particiones ordenadas creadas durante la fase de ordenación de la operación de limpieza. |
| merge_increments | bigint | Cantidad de incrementos de fusión requeridos para completar la fase de fusión de la operación de limpieza. |
| elapsed_time | bigint | Tiempo de ejecución transcurrido de la operación de limpieza (en microsegundos). |
| row_delta | bigint | Diferencia en la cantidad total de filas de la tabla antes y después de la limpieza. |
| sortedrow_delta | bigint | Diferencia en la cantidad de filas ordenadas de la tabla antes y después de la limpieza. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| block_delta | entero | Diferencia en el conteo de bloques para la tabla antes y después de la limpieza. |
| max_merge_partitions | entero | Esta columna se usa para el análisis de rendimiento y representa la cantidad máxima de particiones que la operación de limpieza pueda procesar para la tabla por iteración de fase de fusión. (La limpieza ordena la región desordenada en una o más particiones ordenadas. En función de la cantidad de columnas en la tabla y la configuración actual de Amazon Redshift, la fase de fusión puede procesar una cantidad máxima de particiones en una única iteración de fusión. La fase de fusión funcionará incluso si la cantidad de particiones ordenadas supera la cantidad máxima de particion es de fusión, pero se requerirán más iteraciones de fusión). |

## Consulta de ejemplo

La siguiente consulta devuelve estadísticas para operaciones de limpieza sobre tres tablas diferentes. La tabla SALES se limpió dos veces.

```
select table_name, xid, sort_partitions as parts, merge_increments as merges,
elapsed_time, row_delta, sortedrow_delta as sorted_delta, block_delta
from svv_vacuum_summary
order by xid;

table_  | xid  |parts|merges| elapsed_ | row_    | sorted_ | block_
name    |      |     |      | time     | delta   | delta   | delta
--------+------+-----+------+----------+---------+---------+--------
users   | 2985 |   1 |    1 | 61919653 |       0 |   49990 |      20
category| 3982 |   1 |    1 | 24136484 |       0 |      11 |       0
sales   | 3992 |   2 |    1 | 71736163 |       0 | 1207192 |      32
sales   | 4000 |   1 |    1 | 15363010 | -851648 | -851648 |    -140
(4 rows)
```

# Vistas de monitoreo de SYS

Las vistas de monitoreo son vistas del sistema en Amazon Redshift que se utilizan para monitorizar el uso de recursos de consulta y carga de trabajo de los clústeres y grupos de trabajo sin servidor. Estas vistas se encuentran en el esquema `pg_catalog`. Para mostrar la información que proporcionan estas vistas, ejecute las instrucciones SELECT de SQL.

A menos que se indique lo contrario, estas vistas están disponibles para clústeres de Amazon Redshift y grupos de trabajo de Amazon Redshift sin servidor.

SYS_SERVERLESS_USAGE recopila datos de uso de Amazon Redshift sin servidor únicamente.

Temas


- SYS_ANALYZE_COMPRESSION_HISTORY
- SYS_ANALYZE_HISTORY
- SYS_APPLIED_MASKING_POLICY_LOG
- SYS_AUTO_TABLE_OPTIMIZATION
- SYS_CHILD_QUERY_TEXT
- SYS_CONNECTION_LOG
- SYS_COPY_JOB
- SYS_COPY_JOB_DETAIL
- SYS_COPY_JOB_INFO
- SYS_COPY_REPLACEMENTS
- SYS_DATASHARE_CHANGE_LOG
- SYS_DATASHARE_CROSS_REGION_USAGE
- SYS_DATASHARE_USAGE_CONSUMER
- SYS_DATASHARE_USAGE_PRODUCER
- SYS_EXTERNAL_QUERY_DETAIL
- SYS_EXTERNAL_QUERY_ERROR
- SYS_INTEGRATION_ACTIVITY
- SYS_INTEGRATION_TABLE_ACTIVITY
- SYS_INTEGRATION_TABLE_STATE_CHANGE

- SYS_LOAD_DETAIL
- SYS_LOAD_ERROR_DETAIL
- SYS_LOAD_HISTORY
- SYS_MV_REFRESH_HISTORY
- SYS_MV_STATE
- SYS_PROCEDURE_CALL
- SYS_PROCEDURE_MESSAGES
- SYS_QUERY_DETAIL
- SYS_QUERY_EXPLAIN
- SYS_QUERY_HISTORY
- SYS_QUERY_TEXT
- SYS_RESTORE_LOG
- SYS_RESTORE_STATE
- SYS_SCHEMA_QUOTA_VIOLATIONS
- SYS_SERVERLESS_USAGE
- SYS_SESSION_HISTORY
- SYS_SPATIAL_SIMPLIFY
- SYS_STREAM_SCAN_ERRORS
- SYS_STREAM_SCAN_STATES
- SYS_TRANSACTION_HISTORY
- SYS_UDF_LOG
- SYS_UNLOAD_DETAIL
- SYS_UNLOAD_HISTORY
- SYS_USERLOG
- SYS_VACUUM_HISTORY

## SYS_ANALYZE_COMPRESSION_HISTORY

Registra detalles para las operaciones de análisis de compresión durante los comandos COPY o ANALYZE COMPRESSION.

SYS_ANALYZE_COMPRESSION_HISTORY es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | ID del usuario que generó la entrada. |
| start_time | marca de tiempo | El momento en que la operación de análisis de compresión se inició. |
| transaction_id | bigint | La ID de transacción de la operación de análisis de compresión. |
| table_id | entero | La ID de tabla de la tabla que se analizó. |
| table_name | character (128) | El nombre de la tabala que se analizó. |
| column_position | entero | El índice de la columna de la tabla que se analizó para determinar la codificación de la compresión. |
| old_encoding | character (15) | El tipo de codificación antes del análisis de compresión. |
| new_encoding | character (15) | El tipo de codificación después del análisis de compresión. |
| mode | Carácter(14) | Los valores posibles son:<br>PRESET<br>Especifica que `new_encoding` está determinado por el comando COPY de Amazon Redshift según el tipo de datos de la columna. No se realiza un muestreo de los datos. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | ON<br><br>Especifica que `new_encoding` está determinado por el comando COPY de Amazon Redshift según un análisis de datos de muestra.<br><br>ANALYZE ONLY<br><br>Especifica que `new_encoding` está determinado por el comando ANALYZE COMPRESSION de Amazon Redshift según un análisis de datos de muestra. No obstante, el tipo de codificación de la columna analizada no ha cambiado. |

## Consultas de ejemplo

El siguiente ejemplo inspecciona los detalles del análisis de compresión en la tabla `lineitem` por el último comando COPY ejecutado en la misma sesión.

```
select transaction_id, table_id, btrim(table_name) as table_name, column_position,
 old_encoding, new_encoding, mode
from sys_analyze_compression_history
where transaction_id = (select transaction_id from sys_query_history where query_id =
 pg_last_copy_id()) order by column_position;

 transaction_id  |  table_id   | table_name | column_position |  old_encoding   |
 new_encoding   |      mode
-----------------+-------------+------------+-----------------+-----------------
+-----------------+-------------
      8196       |   248126    | lineitem   |        0        | mostly32        |
 mostly32        | ON
      8196       |   248126    | lineitem   |        1        | mostly32        | lzo
           | ON
      8196       |   248126    | lineitem   |        2        | lzo             |
 delta32k        | ON
      8196       |   248126    | lineitem   |        3        | delta           | delta
           | ON
```

```
      8196       |   248126    | lineitem   |        4        | bytedict        |
 bytedict        | ON
      8196       |   248126    | lineitem   |        5        | mostly32        |
 mostly32        | ON
      8196       |   248126    | lineitem   |        6        | delta           | delta
           | ON
      8196       |   248126    | lineitem   |        7        | delta           | delta
           | ON
      8196       |   248126    | lineitem   |        8        | lzo             | zstd
           | ON
      8196       |   248126    | lineitem   |        9        | runlength       | zstd
           | ON
      8196       |   248126    | lineitem   |       10        | delta           | lzo
           | ON
      8196       |   248126    | lineitem   |       11        | delta           | delta
           | ON
      8196       |   248126    | lineitem   |       12        | delta           | delta
           | ON
      8196       |   248126    | lineitem   |       13        | bytedict        | zstd
           | ON
      8196       |   248126    | lineitem   |       14        | bytedict        | zstd
           | ON
      8196       |   248126    | lineitem   |       15        | text255         | zstd
           | ON
(16 rows)
```

# SYS_ANALYZE_HISTORY

Registra los detalles de las operaciones ANALYZE.

SYS_ANALYZE_HISTORY solo está visible para los superusuarios. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | ID del usuario que generó la entrada. |
| transaction_id | long | El ID de la transacción. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| database_name | char (30) | El nombre de la base de datos. |
| table_name | char (30) | El nombre de la tabla. |
| table_id | entero | ID de la tabla. |
| is_automatic | char(1) | El valor es true (t) si la operación incluía una operación ANALYZE de Amazon Redshift de manera predeterminada. El valor es false (f) si el comando ANALYZE se ejecutaba explícitamente. |
| status | char (15) | El resultado del comando analyze. Los valores posibles son: Full, Skipped y PredicateColumn. |
| start_time | marca de tiempo | La hora en UTC en que comenzó a ejecutarse la operación ANALYZE. |
| end_time | marca de tiempo | La hora en UTC en que terminó de ejecutarse la operación ANALYZE. |
| rows | double | Número total de filas de la tabla |
| modified_rows | double | La cantidad total de filas que se modificaron desde la última operación ANALYZE. |
| analyze_threshold_percent | entero | valor del parámetro analizar_umbral_porcentaje. |
| last_analyze_time | marca de tiempo | Hora en UTC de cuando se analizó previamente la tabla. |

## Consultas de ejemplo

```
 user_id | transaction_id | database_name | schema_name |      table_name     |
 table_id | is_automatic | Status |         start_time         |          end_time
     | rows | modified_rows | analyze_threshold_percent |  last_analyze_time
---------+----------------+---------------+------------+--------------------
+----------+--------------+--------+---------------------------+----------
+-----------------+------+--------------+--------------------------
+---------------------
     101 |           8006 |           dev |      public | test_table_562bf8dc
 |   110427 |            f |   Full | 2023-09-21 18:33:08.504646 | 2023-09-21
 18:33:24.296498 |    5 |             5 |                          0 | 2000-01-01
 00:00:00
```

# SYS_APPLIED_MASKING_POLICY_LOG

Utilice SYS_APPLIED_MASKING_POLICY_LOG para rastrear la aplicación de políticas de enmascaramiento de datos dinámicos en consultas que hacen referencia a relaciones protegidas por DDM.

SYS_APPLIED_MASKING_POLICY_LOG es visible para los usuarios siguientes:

- Superusuarios
- Usuarios con el rol `sys:operator`
- Usuarios con el permiso ACCESS SYSTEM TABLE

Los usuarios normales verán 0 filas.

Tenga en cuenta que SYS_APPLIED_MASKING_POLICY_LOG no es visible para los usuarios con el rol `sys:secadmin`.

Para obtener más información sobre el enmascaramiento de datos dinámico, vaya a Enmascaramiento de datos dinámico.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| policy_name | texto | Nombre de la política de enmascaramiento. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | texto | El ID del usuario que ejecutó la consulta. |
| record_time | marca de tiempo | La hora en la que se registró la entrada de visualización del sistema. |
| session_id | int | El ID del proceso. |
| transaction_id | long | El ID de la transacción. |
| query_id | int | El ID de la consulta. |
| database_name | texto | El nombre de la base de datos en la que se ejecutó la consulta. |
| relation_name | texto | El nombre de la tabla a la que se aplica la política de enmascaramiento. |
| schema_name | texto | El nombre del esquema en el que se encuentra la tabla. |
| attachment_id | long | El ID de la política de enmascaramiento adjunta. |
| relation_kind | texto | El tipo de relación a la que se aplica la política de enmascaramiento. Los valores posibles son `TABLE`, `VIEW`, `LATE BINDING VIEW` y `MATERIALIZED VIEW` . |

## Consultas de ejemplo

El siguiente ejemplo muestra que la política de enmascaramiento `mask_credit_card_full` está adjunta a la tabla `credit_db.public.credit_cards`.

```
select policy_name, database_name, relation_name, schema_name, relation_kind
from sys_applied_masking_policy_log;

policy_name           | database_name | relation_name | schema_name | relation_kind
```

```
----------------------+--------------+---------------+-------------+---------------
mask_credit_card_full | credit_db    | credit_cards  | public      | table

(1 row)
```

# SYS_AUTO_TABLE_OPTIMIZATION

Registra las acciones automatizadas realizadas por Amazon Redshift en tablas definidas para la optimización automática.

SYS_AUTO_TABLE_OPTIMIZATION solo es visible para los superusuarios. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| transaction_id | long | Identificador de la transacción. |
| session_id | int | El identificador de sesión del proceso que ejecutó el comando alter. |
| table_id | int | El identificador de tabla. |
| alter_table_type | character (32) | El tipo de recomendación. Los posibles valores son `distkey`, `sortkey` y `encode`. |
| status | character (128) | El estado de realización de la recomendación. Los valores posibles son `Start`, `Complete`, `Skipped`, `Abort`, `Checkpoint` y `Failed`. |
| event_time | marca de tiempo | La marca temporal de la columna de estado. |
| alter_from | character (200) | El estilo de distribución anterior y las claves de ordenación de la tabla antes de aplicar la recomendación. El valor se trunca en incrementos de 200 caracteres. |

## Consultas de ejemplo

En el siguiente ejemplo, las filas del resultado muestran las acciones realizadas por Amazon Redshift.

```
SELECT table_id, alter_table_type, status, event_time, alter_from
FROM SYS_AUTO_TABLE_OPTIMIZATION;

 table_id |  alter_table_type   |                        status
 |         event_time          |  alter_from
----------+---------------------+-----------------------------------------------------
+----------------------------+-----------------
   118082 | sortkey             | Start
 | 2020-08-22 19:42:20.727049  |
   118078 | sortkey             | Start
 | 2020-08-22 19:43:54.728819  |
   118082 | sortkey             | Start
 | 2020-08-22 19:42:52.690264  |
   118072 | sortkey             | Start
 | 2020-08-22 19:44:14.793572  |
   118082 | sortkey             | Failed
 | 2020-08-22 19:42:20.728917  |
   118078 | sortkey             | Complete
 | 2020-08-22 19:43:54.792705  |  SORTKEY: None;
   118086 | sortkey             | Complete
 | 2020-08-22 19:42:00.72635   |  SORTKEY: None;
   118082 | sortkey             | Complete
 | 2020-08-22 19:43:34.728144  |  SORTKEY: None;
   118072 | sortkey             | Skipped:Retry exceeds the maximum limit for a table.
 | 2020-08-22 19:44:46.706155  |
   118086 | sortkey             | Start
 | 2020-08-22 19:42:00.685255  |
   118082 | sortkey             | Start
 | 2020-08-22 19:43:34.69531   |
   118072 | sortkey             | Start
 | 2020-08-22 19:44:46.703331  |
   118082 | sortkey             | Checkpoint: progress 14.755079%
 | 2020-08-22 19:42:52.692828  |
   118072 | sortkey             | Failed
 | 2020-08-22 19:44:14.796071  |
   116723 | sortkey             | Abort:This table is not AUTO.
 | 2020-10-28 05:12:58.479233  |
```

```
  110203 | distkey             | Abort:This table is not AUTO.
| 2020-10-28 05:45:54.67259   |
```

# SYS_CHILD_QUERY_TEXT

Devuelve el texto de SQL de una consulta secundaria.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | El identificador del usuario que envió la consulta. |
| query_id | bigint | ID de consulta del usuario |
| child_query_sequence | entero | La secuencia de la consulta reescrita del usuario, a partir de 1. |
| sequence | entero | Número de secuencia de este fragmento de consulta. |
| texto | character (200) | Primeros 200 caracteres del texto de la consulta de SQL. |

## Consultas de ejemplo

En el siguiente ejemplo, las filas del resultado muestran las acciones realizadas por Amazon Redshift.

```
SELECT * from sys_child_query_text where query_id = '34487366' order by
 child_query_sequence asc, sequence asc;

user_id | query_id | child_query_sequence | sequence | text
--------|----------|----------------------|----------|-----------------------------------------
100     | 34899339 |   1                  |    0     |  /* RQEV2-aY6ZZ1ZpQK */\nwith
 venue as (\n    select venueid,\n            venuename,\n            venuestate\n
 from venue\n), event as (\n    select eventid,\n            venueid,\n            date
```

```
100     | 34899339 |   1                 |    1     |  id,\n            eventname\n
  from event\n    where eventname like '3 Doors Down'\n), users as (\n    select userid
\n    from users\n), sales as (\n    select salesid,\n            pricepaid,
100     | 34899339 |   1                 |    2     |  \n            eventid,\n
       buyerid\n    from sales\n)\nselect e.eventname,\n        v.venuename,\n
 count(distinct(u.userid)) as unique_customers,\n        sum(s.pricepaid) as total_sal

100     | 34899339 |   1                 |    3     |  es\nfrom venue as v inner join
 event e on v.venueid = e.venueid\ninner join sales s on e.eventid = s.eventid inner
 join users u on s.buyerid = u.userid\ngroup by 1,2\norder by 4 desc limit 100
```

# SYS_CONNECTION_LOG

Registra los intentos de autenticación y las conexiones y desconexiones.

SYS_CONNECTION_LOG solo es visible para los superusuarios. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| event | character(50) | Evento de conexión o de autenticación. |
| record_time | marca de tiempo | Hora cuando se produjo el evento. |
| remote_host | character(45) | Nombre o dirección IP del host remoto. |
| remote_port | character (32) | Número del puerto para el host remoto. |
| session_id | entero | ID del proceso asociado a la instrucción. |
| database_name | character (50) | Nombre de la base de datos. |
| user_name | character (50) | Nombre de usuario. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| auth_method | character (32) | Método de autenticación. |
| duration | entero | Duración de la conexión en microsegundos. |
| ssl_version | character (50) | Versión de la capa de conexión segura (SSL). Esta columna está vacía cuando SSL está desactivado. |
| ssl_cipher | character (128) | Cifrado de la SSL. |
| mtu | entero | Unidad de transmisión máxima (MTU). |
| ssl_compression | character(64) | Tipo de compresión de la SSL. |
| ssl_expansion | character(64) | Tipo de expansión de la SSL. |
| iam_auth_guid | character (36) | ID de autenticación de IAM para la solicitud de CloudTrail. |
| application_name | character (250) | Nombre original o actualizado de la aplicación para una sesión. |
| driver_version | character(64) | La versión del controlador ODBC o JDBC que se conecta a su clúster de Amazon Redshift desde las herramientas de cliente SQL de terceros. |
| os_version | character(64) | La versión del sistema operativo que se encuentra en el equipo cliente que se conecta a su clúster de Amazon Redshift. |
| plugin_name | character (32) | El nombre del complemento utilizado para conectarse a su clúster de Amazon Redshift. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| protocol_version | entero | La versión del protocolo interno que utiliza el controlador de Amazon Redshift cuando se establece su conexión con el servidor. Las versiones de protocolo se negocian entre el controlador y el servidor. En la versión se describen las características disponibles. Los valores válidos son:<br>• 0 (BASE_SERVER_PROTOCOL_VERSION)<br>• 1 (EXTENDED_RESULT_METADATA_SERVER_PROTOCOL_VERSION): para guardar un viaje de ida y vuelta por consulta, el servidor envía información adicional de metadatos del conjunto de resultados.<br>• 2 (BINARY_PROTOCOL_VERSION): según el tipo de datos del conjunto de resultados, el servidor envía datos en formato binario.<br>• 3 (EXTENDED2_RESULT_METADATA_SERVER_PROTOCOL_VERSION): el servidor envía información (de intercalación) con distinción de mayúsculas y minúsculas de una columna. |
| global_session_id | character (36) | El identificador único global de la sesión actual. El ID de la sesión persiste durante los reinicios de errores de nodo. |

## Consultas de ejemplo

Para ver los detalles de las conexiones abiertas, ejecute la siguiente consulta.

```
select record_time, user_name, database_name, remote_host, remote_port
from sys_connection_log
where event = 'initiating session'
and session_id not in
(select session_id from sys_connection_log
where event = 'disconnecting session')
order by 1 desc;
```

```
record_time         | user_name   | database_name   | remote_host   | remote_port

--------------------+-------------+-----------------+---------------
+--------------------------------
2014-11-06 20:30:06 | rdsdb       | dev             | [local]       |

2014-11-06 20:29:37 | test001     | test            | 10.49.42.138  | 11111

2014-11-05 20:30:29 | rdsdb       | dev             | 10.49.42.138  | 33333

2014-11-05 20:28:35 | rdsdb       | dev             | [local]       |
(4 rows)
```

En el siguiente ejemplo, se refleja un intento fallido de autenticación y una correcta conexión y desconexión.

```
select event, record_time, remote_host, user_name
from sys_connection_log order by record_time;

            event      |         record_time        |  remote_host  | user_name

-----------------------+----------------------------+---------------+---------
authentication failure | 2012-10-25 14:41:56.96391  | 10.49.42.138  | john

authenticated          | 2012-10-25 14:42:10.87613  | 10.49.42.138  | john

initiating session     | 2012-10-25 14:42:10.87638  | 10.49.42.138  | john

disconnecting session  | 2012-10-25 14:42:19.95992  | 10.49.42.138  | john

(4 rows)
```

En el siguiente ejemplo, se muestra la versión del controlador ODBC, el sistema operativo del equipo cliente y el complemento utilizado para conectarse al clúster de Amazon Redshift. En este ejemplo, el complemento se utiliza para la autenticación de controladores ODBC estándar mediante un nombre de inicio de sesión y una contraseña.

```
select driver_version, os_version, plugin_name from sys_connection_log;

driver_version                          |  os_version                       |
 plugin_name
```

```
----------------------------------------+-----------------------------------
+--------------------
Amazon Redshift ODBC Driver 1.4.15.0001 | Darwin 18.7.0 x86_64               | none
Amazon Redshift ODBC Driver 1.4.15.0001 | Linux 4.15.0-101-generic x86_64    | none
```

En el siguiente ejemplo, se muestra la versión del sistema operativo en el equipo cliente, la versión del controlador y la versión del protocolo.

```
select os_version, driver_version, protocol_version from sys_connection_log;

os_version                      |  driver_version              | protocol_version
--------------------------------+------------------------------+--------------------
Linux 4.15.0-101-generic x86_64 | Redshift JDBC Driver 2.0.0.0 | 2
Linux 4.15.0-101-generic x86_64 | Redshift JDBC Driver 2.0.0.0 | 2
Linux 4.15.0-101-generic x86_64 | Redshift JDBC Driver 2.0.0.0 | 2
```

# SYS_COPY_JOB

Utilice SYS_COPY_JOB para ver los detalles de los comandos COPY JOB.

Esta vista contiene los comandos COPY JOB que se han creado.

SYS_COPY_JOB es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| job_id | bigint | Identificador del trabajo de copia. |
| job_name | character (128) | Nombre del trabajo de copia. |
| iam_role | character (128) | Rol de IAM especificado en la instrucción COPY. |
| job_text | character (256) | Parámetros de la instrucción COPY. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| is_auto | entero | Indica si Amazon Redshift ejecuta automáticamente COPY JOB. Un 1 indica true y un 0 indica false. |
| on_error_suspend | entero | Esta información es solo para uso interno. |

## SYS_COPY_JOB_DETAIL

Utilice SYS_COPY_JOB_DETAIL para ver los detalles de los comandos COPY JOB.

Esta vista contiene los comandos COPY JOB que se han creado. Si COPY JOB intenta cargar un archivo y este no se carga, el archivo se omitirá en futuros intentos automáticos de COPY JOB.

SYS_COPY_JOB_DETAIL es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

### Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | Identificador del usuario propietario del trabajo. |
| database_name | texto | Nombre de la base de datos que contiene el trabajo. |
| job_id | entero | Identificador del trabajo de copia. |
| job_name | texto | Nombre del trabajo de copia. |
| file_location | texto | Nombre del bucket de Amazon S3 para la entrada. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| file_name | texto | Nombre de la clave de Amazon S3 para la entrada. |
| file_size | bigint | Tamaño de la entrada, en bytes. |
| file_etag | texto | Etiqueta de entidad de Amazon S3 (ETag) de la entrada. |
| job_text | character (256) | Parámetros de la instrucción COPY. |
| modification_time | marca de tiempo | Hora (UTC) a la que se creó o modificó el objeto de S3 al iniciar la ingesta. En el modo de enumeración de S3, es la hora de modificación del objeto. En el modo de notificación, este es el momento en que se produjo el evento de notificación. |
| enqueue_time | marca de tiempo | Hora (UTC) a la que se añadió la entrada a la cola de ingesta. |
| status | char(1) | Estado del evento de la entrada. Valores válidos: `P` para ingesta pendiente, `I` para ingeridos, `E` para la ingesta de errores y `U` para desconocido (lo que indica algún tipo de estado inesperado). |

El ejemplo siguiente devuelve una fila donde se muestran las entradas ingeridas.

```
SELECT * FROM SYS_COPY_JOB_DETAIL WHERE status ilike '%ingested%' limit 1;


user_id | 100
database_name | dev
job_name | many_job_4_3
job_id | 110702
file_location | saral-sqs-system4623202051-0
file_name | frenzy-9/4623202051/file_0_107
file_size | 11302
file_etag | 51b2d78ac5b5aecf4ee6f8374815ad19
modification_time | 2024-07-15 20:43:14
enqueue_time | 2024-07-15 20:44:24
status | Ingested
```

# SYS_COPY_JOB_INFO

Utilice SYS_COPY_JOB_INFO para visualizar los mensajes registrados sobre un COPY JOB.

Esta vista contiene información sobre los errores de un COPY JOB que se ha ejecutado.

SYS_COPY_JOB_INFO es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| job_id | bigint | Identificador del trabajo de copia. |
| job_name | character (128) | Nombre del trabajo de copia. |
| database_name | character (128) | El nombre de la base de datos. |
| record_time | marca de tiempo | Hora (UTC) a la que se registró el mensaje de error. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| message | character(512) | Este mensaje es del evento registrado para un COPY JOB. |

## SYS_COPY_REPLACEMENTS

Muestra un registro de los momentos en que se reemplazaron caracteres UTF-8 no válidos por el comando COPY con la opción ACCEPTINVCHARS. Se agrega una entrada de registro a SYS_COPY_REPLACEMENTS por cada una de las 100 primeras filas de cada sector del nodo que necesitaron al menos un reemplazo.

Puede utilizar esta vista para consultar información sobre los grupos de trabajo sin servidor y los clústeres aprovisionados.

SYS_COPY_REPLACEMENTS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

### Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | ID del usuario que generó la consulta. |
| query_id | bigint | El ID de la consulta. Columna de consulta usada para combinar otras vistas y tablas del sistema. |
| table_id | entero | El ID de la tabla. |
| file_name | character (256) | Ruta completa al archivo de entrada para el comando COPY. |
| column_name | character (127) | Primer campo que tiene un carácter UTF-8 no válido. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| line_number | bigint | Número de línea del archivo de datos de entrada que contiene un carácter UTF-8 no válido. El valor -1 indica que el número de línea no está disponible, por ejemplo, al copiar desde un archivo de datos en columnas. |
| raw_line | character (1024) | Datos de carga sin procesar que tienen un carácter UTF-8 no válido. |

### Consultas de ejemplo

El siguiente ejemplo devuelve los reemplazos de la operación COPY más reciente.

```
select query_idp, table_id, file_name, line_number, colname
from sys_copy_replacements
where query = pg_last_copy_id();


 query_id | table_id |   file_name                                                |
 line_number | column_name
 ---------+----------+------------------------------------------------------------
+-------------+--------
    96    |    26    | s3://DOC-EXAMPLE-BUCKET/allusers_pipe.txt             |
 123 | city
    96    |    26    | s3://DOC-EXAMPLE-BUCKET/allusers_pipe.txt             |
 456 | city
    96    |    26    | s3://DOC-EXAMPLE-BUCKET/allusers_pipe.txt             |
 789 | city
    96    |    26    | s3://DOC-EXAMPLE-BUCKET/allusers_pipe.txt             |
 012 | city
    96    |    26    | s3://DOC-EXAMPLE-BUCKET/allusers_pipe.txt             |
 119 | city
...
```

## SYS_DATASHARE_CHANGE_LOG

Registra la vista consolidada para realizar el seguimiento de los cambios en los datashares (recursos para compartir datos) tanto en los clústeres productores como en los consumidores.

SYS_DATASHARE_CHANGE_LOG es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | El ID del grupo de usuarios que realizan la acción. |
| user_name | varchar(128) | El nombre del usuario que realiza la acción. |
| session_id | entero | El ID de la sesión. |
| transaction_id | bigint | El ID de la transacción. |
| share_id | entero | El ID del datashare afectado. |
| share_name | varchar(128) | El nombre del datashare. |
| source_database_id | entero | El ID de la base de datos a la que pertenece el datashare. |
| source_database_name | varchar(128) | El nombre de la base de datos a la que pertenece el datashare. |
| consumer_database_id | entero | El ID de la base de datos importada desde el datashare. |
| consumer_database_name | varchar(128) | El nombre de la base de datos importada desde el datashare. |
| arn | varchar(192) | El ARN del recurso que respalda la base de datos importada. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| record_time | marca de tiempo | La marca temporal de la acción. |
| acción | varchar(128) | La acción que se está ejecutando. Los valores posibles son CREATE DATASHARE, DROP DATASHARE, GRANT ALTER, REVOKE ALTER, GRANT SHARE, REVOKE SHARE, ALTER ADD, ALTER REMOVE, ALTER SET, GRANT USAGE, REVOKE USAGE, CREATE DATABASE, GRANT o REVOKE USAGE en una base de datos compartida, DROP SHARED DATABASE, ALTER SHARED DATABASE. |
| status | entero | El estado de la acción. Los valores posibles son SUCCESS y ERROR-ERROR CODE. |
| share_object_type | varchar(64) | El tipo de objeto de base de datos que se agregó al datashare o se eliminó de él. Los valores posibles son esquemas, tablas, columnas, funciones y vistas. Este es un campo para el clúster productor. |
| share_object_id | entero | El ID del objeto de base de datos que se agregó al datashare o se eliminó de él. Este es un campo para el clúster productor. |
| share_object_name | varchar(128) | El nombre del objeto de base de datos que se agregó al datashare o se eliminó de él. Este es un campo para el clúster productor. |
| target_user_type | varchar(16) | El tipo de usuarios o grupos a los que se concedió un privilegio. Este es un campo tanto para el clúster productor como para el consumidor. |
| target_user_id | entero | Los ID de usuarios o grupos a los que se concedió un privilegio. Este es un campo tanto para el clúster productor como para el consumidor. |
| target_user_name | varchar(128) | El nombre de los usuarios o grupos a los que se concedió un privilegio. Este es un campo tanto para el clúster productor como para el consumidor. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| consumer_account | varchar (16) | El ID de cuenta del consumidor de datos. Este es un campo para el clúster productor. |
| consumer_namespace | varchar (64) | El espacio de nombres de la cuenta consumidora de datos. Este es un campo para el clúster productor. |
| producer_account | varchar (16) | El ID de la cuenta productora a la que pertenece el datashare. Este es un campo para el clúster consumidor. |
| producer_namespace | varchar (64) | El espacio de nombres de la cuenta de producto a la que pertenece el datashare. Este es un campo para el clúster consumidor. |
| attribute_name | varchar (64) | El nombre de un atributo de datashare o base de datos compartida. |
| attribute_value | varchar(128) | El valor de un atributo de datashare o base de datos compartida. |
| message | varchar(512) | El mensaje de error cuando se produce un error en una acción. |

## Consultas de ejemplo

En el siguiente ejemplo, se muestra una vista SYS_DATASHARE_CHANGE_LOG.

```
SELECT DISTINCT action
FROM sys_datashare_change_log
WHERE share_object_name LIKE 'tickit%';

         action
 ----------------------
  "ALTER DATASHARE ADD"
```

# SYS_DATASHARE_CROSS_REGION_USAGE

Utilice la vista SYS_DATASHARE_CROSS_REGION_USAGE para obtener un resumen del uso de datos transferidos entre regiones debido a la consulta de uso compartido de datos entre regiones. SYS_DATASHARE_CROSS_REGION_USAGE agrega detalles en el nivel de segmento.

Solo los superusuarios pueden ver SYS_DATASHARE_CROSS_REGION_USAGE. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| query_id | entero | El Id. de la consulta. Utilice este valor para unir otras tablas y vistas del sistema. |
| child_query_sequence | entero | La secuencia de la consulta reescrita del usuario, a partir de 1. |
| segment_id | bigint | Número del segmento. Una consulta consta de distintos segmentos y cada segmento consta de uno o más pasos. |
| start_time | hora | Es la hora en UTC en que se ha iniciado la transferencia de datos. |
| end_time | hora | Es la hora en UTC en que ha finalizado la transferencia de datos. |
| transferred_data | bigint | Número de bytes de datos transferidos de una región productora a una región consumidora. |
| source_region | char(25) | Región productora desde la que la consulta transfirió los datos. |

## Consultas de ejemplo

En el siguiente ejemplo, se muestra una vista SYS_DATASHARE_CROSS_REGION_USAGE.

```
SELECT query_id, segment_id, transferred_data, source_region
from sys_datashare_cross_region_usage
```

```
where query_id = pg_last_query_id()
order by query_id, segment_id;

  query_id | segment_id | transferred_data | source_region
-----------+------------+------------------+---------------
    200048 |          2 |          4194304 |    us-west-1
    200048 |          2 |          4194304 |    us-east-2
```

# SYS_DATASHARE_USAGE_CONSUMER

Registra la actividad y el uso de los datashares. Esta vista solo es relevante en el clúster consumidor.

SYS_DATASHARE_USAGE_CONSUMER es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | El ID del usuario que emite la solicitud. |
| session_id | entero | El ID del proceso principal que ejecuta la consulta. |
| transaction_id | bigint | El contexto de la transacción actual. |
| request_id | varchar(50) | El ID único de la llamada a la API solicitada. |
| request_type | varchar(25) | El tipo de la solicitud realizada al clúster productor. |
| transaction_uid | varchar(50) | El ID único de la transacción. |
| record_time | marca de tiempo | La hora en que se registra la acción. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| status | entero | El estado de la llamada a la API solicitada. |
| error_message | varchar(512) | El mensaje de un error. |

### Consultas de ejemplo

En el siguiente ejemplo, se muestra la vista SYS_DATASHARE_USAGE_CONSUMER.

```
SELECT request_type, status, trim(error) AS error
FROM sys_datashare_usage_consumer

  request_type  | status | error_message
----------------+--------+---------------
 "GET RELATION" |   0    |
```

## SYS_DATASHARE_USAGE_PRODUCER

Registra la actividad y el uso de los datashares. Esta vista solo es relevante en el clúster productor.

SVL_DATASHARE_USAGE_PRODUCER es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

### Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| share_id | entero | El ID de objeto (OID) del datashare. |
| share_name | varchar(128) | El nombre del datashare. |
| request_id | varchar(50) | El ID único de la llamada a la API solicitada. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| request_type | varchar(25) | El tipo de la solicitud realizada al clúster productor. |
| object_type | varchar (64) | El tipo de objeto que se comparte desde el datashare. Los valores posibles son esquemas, tablas, columnas, funciones y vistas. |
| object_oid | entero | El ID del objeto que se comparte desde el datashare. |
| object_name | varchar(128) | El nombre del objeto que se comparte desde el datashare. |
| consumer_account | varchar (16) | La cuenta de la cuenta consumidora con la que se comparte el datashare. |
| consumer_namespace | varchar (64) | El espacio de nombres de la cuenta consumidora con la que se comparte el datashare. |
| consumer_transaction_uid | varchar(50) | El ID de transacción único de la instrucción en el clúster consumidor. |
| record_time | marca de tiempo | La hora en que se registra la acción. |
| status | entero | El estado del datashare. |
| error_message | varchar(512) | El mensaje de un error. |
| consumer_region | char(64) | Es la región en la que se encuentra el clúster del consumidor. |

## Consultas de ejemplo

En el siguiente ejemplo, se muestra la vista SYS_DATASHARE_USAGE_PRODUCER.

```
SELECT DISTINCT
FROM sys_datashare_usage_producer
WHERE object_name LIKE 'tickit%';

   request_type
 ------------------
   "GET RELATION"
```

# SYS_EXTERNAL_QUERY_DETAIL

Utilice SYS_EXTERNAL_QUERY_DETAIL para visualizar los detalles de las consultas en el nivel del segmento. Cada fila representa un segmento de una determinada consulta de WLM con una serie de detalles como el número de filas procesadas, el número de bytes procesados y la información de partición de las tablas externas en Amazon S3. Cada fila de esta vista tendrá también una entrada correspondiente en la vista SYS_QUERY_DETAIL, con la diferencia de que esta última contiene más información detallada relativa al procesamiento de las consultas externas.

SYS_EXTERNAL_QUERY_DETAIL es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | El identificador del usuario que envió la consulta. |
| query_id | bigint | El identificador de consulta de la consulta externa. |
| transaction_id | bigint | Identificador de la transacción. |
| child_query_sequence | entero | La secuencia de la consulta reescrita del usuario. Comienza por 0, parecido a segment_id. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| segment_id | entero | El identificador del segmento de la consulta. |
| source_type | character (32) | El tipo de origen de datos de la consulta podría ser S3 para Redshift Spectrum, PG para consultas federadas. |
| start_time | marca de tiempo | La hora de inicio de la consulta. |
| end_time | marca de tiempo | La hora en que se completó la consulta. |
| duración | bigint | La cantidad de tiempo (microsegundos) empleado en la consulta. |
| total_partitions | entero | La cantidad de particiones necesarias para una consulta de Amazon S3. |
| qualified_partitions | entero | La cantidad de particiones que ha analizado una consulta de Amazon S3. |
| scanned_files | bigint | La cantidad de archivos analizados de Amazon S3. |
| returned_rows | bigint | La cantidad de filas analizadas para una consulta de Amazon S3, o la cantidad de filas devueltas para una consulta federada. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| returned_bytes | bigint | La cantidad de bytes analizados para una consulta de Amazon S3, o la cantidad de bytes devueltos para una consulta federada. |
| file_format | texto | El formato de archivos de los archivos de Amazon S3. |
| file_location | texto | Ubicación de Amazon S3 de la tabla externa. |
| external_query_text | texto | El texto de consulta en el nivel del segmento para una consulta federada. |
| warning_message | character (4000) | El mensaje de advertencia que se muestra cuando se ejecuta la consulta. |
| table_name | character (136) | Nombre de la tabla del paso que está en curso. |
| es recursivo | character (1) | Indica si se realiza un análisis recursivo de las subcarpetas. |
| is_nested | character (1) | Indica si se accede al tipo de datos de columna anidada. |
| s3list_time | bigint | Duración del listado de archivos en milisegundos. |
| get_partition_time | long | Tiempo empleado en enumerar y calificar las particiones de un objeto externo dado de AWS Glue Data Catalog y Apache Hive. |

## Consultas de ejemplo

La siguiente consulta muestra los detalles de la consulta externa.

```
SELECT query_id,
       segment_id,
       start_time,
       end_time,
       total_partitions,
       qualified_partitions,
       scanned_files,
       returned_rows,
       returned_bytes,
       trim(external_query_text) query_text,
       trim(file_location) file_location
FROM sys_external_query_detail
ORDER BY query_id, start_time DESC
LIMIT 2;
```

Resultados de ejemplo.

```
 query_id | segment_id |         start_time         |          end_time
   | total_partitions | qualified_partitions | scanned_files | returned_rows |
 returned_bytes | query_text | file_location
----------+------------+----------------------------+----------------------------
+------------------+----------------------+---------------+---------------
+----------------+------------+---------------
   763251 |          0 | 2022-02-15 22:32:23.312448 | 2022-02-15 22:32:24.036023 |
          3 |                    3 |             3 |         38203 |        2683414 |
          |
   763254 |          0 | 2022-02-15 22:32:40.17103  | 2022-02-15 22:32:40.839313 |
          3 |                    3 |             3 |         38203 |        2683414 |
          |
```

# SYS_EXTERNAL_QUERY_ERROR

Puede consultar la vista del sistema SYS_EXTERNAL_QUERY_ERROR para obtener información acerca de los errores de análisis de Redshift Spectrum. SYS_EXTERNAL_QUERY_ERROR muestra un ejemplo de los errores registrados. El valor predeterminado es 10 entradas por consulta.

SYS_EXTERNAL_QUERY_ERROR es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | Identificador del usuario que generó la consulta. |
| query_id | bigint | Identificaodr de la consulta que generó esta fila. |
| file_location | char (256) | Ubicación de los datos que se están consultando. |
| rowid | char(2100) | La ubicación del error en el archivo. Las partes de `rowid` se separan con `:` (dos puntos), y es posible agregar otras partes en el futuro.<br>`row_offset :row_group :row_id`<br>Un row_offset es el desplazamiento (en bytes) de la fila dentro del archivo, y se establece en `-1` para formatos de archivo no admitidos. Una tabla se divide en row_groups, y cada grupo tiene filas con distintos row_ids. |
| column_name | char(127) | Nombre de la columna que ha devuelto la consulta. |
| original_value | char (1024) | Valor original consultado. |
| modified_value | char (1024) | Valor modificado devuelto en función de la opción de configuración de control de datos especificada en la consulta. |
| Desencadenador | char(128) | Opción de control de datos especificada en la consulta. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| acción | char(128) | Acción asociada a la opción de control de datos especific ada en la consulta. |
| action_va lue | char(128) | Valor del parámetro de acción asociado a la opción de control de datos especificada en la consulta. |
| error_code | entero | Código del resultado de la opción de control de datos especificada en la consulta. |

## Consulta de ejemplo

La siguiente consulta devuelve la lista de filas para las que se han realizado operaciones de control de datos.

```
SELECT * FROM sys_external_query_error;
```

La consulta devuelve resultados similares a los siguientes.

```
  user_id   query_id  file_location                                rowid
column_name           original_value             modified_value       trigger
action               action_value                 error_code
    100     1574007   s3://spectrum-uddh/league/spi_global_rankings.0:0
league_name           Barclays Premier League    Barclays Premier Lea UNSPECIFIED
TRUNCATE                                          156
    100     1574007   s3://spectrum-uddh/league/spi_global_rankings.0:0
league_nspi           34595                      32767                UNSPECIFIED
OVERFLOW_VALUE                                    199
    100     1574007   s3://spectrum-uddh/league/spi_global_rankings.0:1
league_nspi           34151                      32767                UNSPECIFIED
OVERFLOW_VALUE                                    199
    100     1574007   s3://spectrum-uddh/league/spi_global_rankings.0:2
league_name           Barclays Premier League    Barclays Premier Lea UNSPECIFIED
TRUNCATE                                          156
    100     1574007   s3://spectrum-uddh/league/spi_global_rankings.0:2
league_nspi           33223                      32767                UNSPECIFIED
OVERFLOW_VALUE                                    199
```

```
    100     1574007   s3://spectrum-uddh/league/spi_global_rankings.0:3
league_name           Barclays Premier League    Barclays Premier Lea UNSPECIFIED
TRUNCATE                                          156
    100     1574007   s3://spectrum-uddh/league/spi_global_rankings.0:3
league_nspi           32808                      32767                UNSPECIFIED
OVERFLOW_VALUE                                    199
    100     1574007   s3://spectrum-uddh/league/spi_global_rankings.0:4
league_nspi           32790                      32767                UNSPECIFIED
OVERFLOW_VALUE                                    199
    100     1574007   s3://spectrum-uddh/league/spi_global_rankings.0:5
league_name           Spanish Primera Division   Spanish Primera Divi UNSPECIFIED
TRUNCATE                                          156
    100     1574007   s3://spectrum-uddh/league/spi_global_rankings.0:6
league_name           Spanish Primera Division   Spanish Primera Divi UNSPECIFIED
TRUNCATE                                          156
```

# SYS_INTEGRATION_ACTIVITY

SYS_INTEGRATION_ACTIVITY muestra detalles sobre las ejecuciones de integración completadas.

SYS_INTEGRATION_ACTIVITY solo está visible para los superusuarios. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Para obtener información sobre las integraciones sin ETL, consulte Uso de integraciones sin ETL en la guía de administración de Amazon Redshift.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| integration_id | character (128) | Identificador asociado con la integración. |
| target_database | character (128) | Base de datos de Amazon Redshift que recibe los datos de integración. |
| origen | character (128) | Datos de origen de la integración. Los tipos posibles son `MySQL` y `PostgreSQL` . |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| checkpoint_name | character (128) | Nombre del punto de control que replica las coordenadas de binlog. |
| checkpoint_type | character (16) | Tipo de punto de control. Entre los valores posibles se incluyen: `snapshot` y `cdc`. |
| checkpoint_bytes | bigint | Número de bytes de este punto de control. |
| last_commit_timestamp | marca de tiempo | Marca temporal de la última confirmación en este punto de control. |
| modified_tables | entero | Número de tablas modificadas en el punto de control. |
| integration_start_time | hora | Hora (UTC) en la que se inició la integración para este punto de control. |
| integration_end_time | hora | Hora (UTC) en la que finalizó la integración para este punto de control. |

## Consultas de ejemplo

El siguiente comando SQL muestra el registro de las integraciones.

```
select * from sys_integration_activity;

          integration_id              | target_database | source |
   checkpoint_name                 | checkpoint_type  | checkpoint_bytes |
 last_commit_timestamp   | modified_tables |   integration_start_time   |
 integration_end_time
-------------------------------------+----------------+--------
+-----------------------------------------------+-----------------+-----------------
+------------------------+----------------+---------------------------
+---------------------------
 76b15917-afae-4447-b7fd-08e2a5acce7b |   demo1         | MySQL  | checkpoints/
checkpoint_3_241_3_510.json     |        cdc       |        762       | 2023-05-10
 23:00:14.201 |         1       | 2023-05-10 23:00:45.054265 | 2023-05-10
 23:00:46.339826
```

```
 76b15917-afae-4447-b7fd-08e2a5acce7b |   demo1         | MySQL  | checkpoints/
checkpoint_3_16329_3_17839.json |        cdc       |       13488      | 2023-05-11
 01:33:57.411 |         2       | 2023-05-11 02:19:09.440121 | 2023-05-11
 02:19:16.090492
 76b15917-afae-4447-b7fd-08e2a5acce7b |   demo1         | MySQL  | checkpoints/
checkpoint_3_5103_3_5532.json   |        cdc       |        1657      | 2023-05-10
 23:13:14.205 |         2       | 2023-05-10 23:13:23.545487 | 2023-05-10
 23:13:25.652144
```

# SYS_INTEGRATION_TABLE_ACTIVITY

SYS_INTEGRATION_TABLE_ACTIVITY muestra los detalles de la actividad de inserción, eliminación y actualización de las integraciones sin ETL. Se añade una fila por cada ingesta completada.

Un superusuario puede ver todas las filas de esta tabla.

Para obtener más información, consulte Integraciones sin ETL.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| integration_id | character (128) | Identificador asociado con la integración. |
| checkpoint_name | character (128) | Nombre del punto de comprobación. |
| target_database | character (128) | Se trata del nombre de la base de datos de Amazon Redshift. |
| schema_name | character (128) | Nombre del esquema de Amazon Redshift. |
| table_name | character (128) | El nombre de la tabla. |
| table_id | entero | Identificador de la tabla. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| record_time | marca de tiempo | Hora (UTC) a la que se realizó el cambio. |
| transaction_id | bigint | Identificador de la transacción. |
| inserted_rows | bigint | Número de filas que ha insertado la ingesta. |
| deleted_rows | bigint | Número de filas que ha eliminado la ingesta. |
| updated_rows | bigint | Número de filas que ha actualizado la ingesta. |
| bytes_ingested | bigint | Número de bytes ingeridos. |

### Consultas de ejemplo

El siguiente comando de SQL muestra la actividad de la integración.

```
select * from sys_integration_table_activity;

          integration_id              | checkpoint_name | target_database | schema_name
 |     table_name    | table_id     | record_time                | transaction_id  |
 inserted_rows  | deleted_rows | updated_rows | bytes_ingested
--------------------------------------+-----------------+-----------------
+-------------+------------------+-------------+---------------------------
+-----------------+----------------+-------------+-------------+---------------
 4798e675-8f9f-4686-b05f-92c538e19629 |                 | sample_test2    |    sample
 | SampleTestChannel |  111276      | 2023-05-12 12:40:30.656625 | 7736            |  2
             | 0            | 0            | 125
```

## SYS_INTEGRATION_TABLE_STATE_CHANGE

SYS_INTEGRATION_TABLE_STATE_CHANGE: muestra los detalles de los registros de cambios de estados de tabla para las integraciones.

Un superusuario puede ver todas las filas de esta tabla.

Para obtener más información, consulte Uso de integraciones sin ETL.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| integration_id | character (128) | Identificador asociado con la integración. |
| database_name | character (128) | Se trata del nombre de la base de datos de Amazon Redshift. |
| schema_name | character (128) | Nombre del esquema de Amazon Redshift. |
| table_name | character (128) | El nombre de la tabla. |
| new_state | character (128) | Estado de la tabla. Los valores posibles son `Synced`, `ResyncRequired` , `ResyncInitiated` , `Deleted`, `Failed` y `ResyncDeleted` . |
| table_last_replicated_checkpoint | character (128) | Coordenadas del registro sincronizado actual. |
| state_change_reason | character (256) | Motivo de la última transición de estado. |
| record_time | marca de tiempo | Hora (UTC) en que se actualizó el registro por última vez. |

## Consultas de ejemplo

El siguiente comando SQL muestra el registro de las integraciones.

```
select * from sys_integration_table_state_change;

           integration_id            | database_name | schema_name | table_name
 | new_state |  table_last_replicated_checkpoint   | state_change_reason |
 record_time
```

```
-------------------------------------+--------------+-------------+------------
+-----------+-------------------------------------+--------------------
+---------------------------
 99108e72-1cfd-414f-8cc0-0216acefac77 | perfdb        | sbtest80t3s | sbtest79   |
 Synced    | {"txn_seq":9834,"txn_id":126597515} |                    | 2023-09-20
 19:39:50.087868
 99108e72-1cfd-414f-8cc0-0216acefac77 | perfdb        | sbtest80t3s | sbtest56   |
 Synced    | {"txn_seq":9834,"txn_id":126597515} |                    | 2023-09-20
 19:39:45.54005
 99108e72-1cfd-414f-8cc0-0216acefac77 | perfdb        | sbtest80t3s | sbtest50   |
 Synced    | {"txn_seq":9834,"txn_id":126597515} |                    | 2023-09-20
 19:40:20.362504
 99108e72-1cfd-414f-8cc0-0216acefac77 | perfdb        | sbtest80t3s | sbtest18   |
 Synced    | {"txn_seq":9834,"txn_id":126597515} |                    | 2023-09-20
 19:40:32.544084
 99108e72-1cfd-414f-8cc0-0216acefac77 | perfdb        | sbtest40t3s | sbtest23   |
 Synced    | {"txn_seq":9834,"txn_id":126597515} |                    | 2023-09-20
 15:49:05.186209
```

# SYS_LOAD_DETAIL

Devuelve información para realizar un seguimiento o solucionar problemas en la carga de datos.

En esta vista, se registra el progreso de cada archivo de datos mientras se carga en una tabla de base de datos.

Todos los usuarios pueden acceder a esta vista. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | ID del usuario que generó la entrada. |
| query_id | entero | ID de la consulta. |
| file_name | character (256) | Nombre del archivo que se va a cargar. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| bytes_scanned | entero | La cantidad de bytes analizados desde el archivo de Amazon S3. |
| lines_scanned | entero | Cantidad de líneas examinadas del archivo de carga. Este número puede no coincidir con el número de filas que se cargan efectivamente. Por ejemplo, la carga puede examinar y, al mismo tiempo, tolerar una serie de registros incorrectos, en función de la opción MAXERROR en el comando COPY. |
| record_time | marca de tiempo | Hora en que se actualizó por última vez esta entrada. |
| splits_scan | Número de divisiones de este archivo. | Número de divisiones de este archivo. |
| start_time | marca de tiempo | Hora a la que se inició el procesamiento de este archivo. |
| end_time | marca de tiempo | Hora a la que finalizó el procesamiento de este archivo. |

## Consultas de ejemplo

El siguiente ejemplo devuelve los detalles de la última operación COPY.

```
select query_id, trim(file_name) as file, record_time
from sys_load_detail
where query_id = pg_last_copy_id();

 query_id |               file                |          record_time
----------+-----------------------------------+----------------------------
 28554    | s3://dw-tickit/category_pipe.txt | 2013-11-01 17:14:52.648486
(1 row)
```

La siguiente consulta tiene entradas para una carga nueva de las tablas en la base de datos TICKIT:

```
select query_id, trim(file_name), record_time
from sys_load_detail
where file_name like '%tickit%' order by query_id;
```

```
 query_id |           btrim           |          record_time
----------+---------------------------+---------------------------
 22475    | tickit/allusers_pipe.txt | 2013-02-08 20:58:23.274186
 22478    | tickit/venue_pipe.txt    | 2013-02-08 20:58:25.070604
 22480    | tickit/category_pipe.txt | 2013-02-08 20:58:27.333472
 22482    | tickit/date2008_pipe.txt | 2013-02-08 20:58:28.608305
 22485    | tickit/allevents_pipe.txt| 2013-02-08 20:58:29.99489
 22487    | tickit/listings_pipe.txt | 2013-02-08 20:58:37.632939
 22593    | tickit/allusers_pipe.txt | 2013-02-08 21:04:08.400491
 22596    | tickit/venue_pipe.txt    | 2013-02-08 21:04:10.056055
 22598    | tickit/category_pipe.txt | 2013-02-08 21:04:11.465049
 22600    | tickit/date2008_pipe.txt | 2013-02-08 21:04:12.461502
 22603    | tickit/allevents_pipe.txt| 2013-02-08 21:04:14.785124
 22605    | tickit/listings_pipe.txt | 2013-02-08 21:04:20.170594

(12 rows)
```

El hecho de que se escriba un registro en el archivo de registros para esta vista de sistema no implica que la carga se haya confirmado correctamente como parte de la transacción que contiene. Para comprobar las confirmaciones de carga, consulte la vista STL_UTILITYTEXT y busque el registro COMMIT que corresponda a esa transacción COPY. Por ejemplo, esta consulta combina SYS_LOAD_DETAIL y STL_QUERY en función de una subconsulta de STL_UTILITYTEXT:

```
select l.query_id,rtrim(l.file_name),q.transaction_id
from sys_load_detail l, sys_query_text q
where l.query_id=q.query_id
and exists
(select xid from stl_utilitytext where xid=q.transaction_id and
 rtrim("text")='COMMIT');

 query_id |           rtrim           |  transaction_id
----------+---------------------------+-----------------
 22600    | tickit/date2008_pipe.txt  | 68311
 22480    | tickit/category_pipe.txt  | 68066
  7508    | allusers_pipe.txt         | 23365
  7552    | category_pipe.txt         | 23415
  7576    | allevents_pipe.txt        | 23429
  7516    | venue_pipe.txt            | 23390
  7604    | listings_pipe.txt         | 23445
 22596    | tickit/venue_pipe.txt     | 68309
 22605    | tickit/listings_pipe.txt  | 68316
 22593    | tickit/allusers_pipe.txt  | 68305
```

```
 22485    | tickit/allevents_pipe.txt | 68071
  7561    | allevents_pipe.txt        | 23429
  7541    | category_pipe.txt         | 23415
  7558    | date2008_pipe.txt         | 23428
 22478    | tickit/venue_pipe.txt     | 68065
   526    | date2008_pipe.txt         |  2572
  7466    | allusers_pipe.txt         | 23365
 22482    | tickit/date2008_pipe.txt  | 68067
 22598    | tickit/category_pipe.txt  | 68310
 22603    | tickit/allevents_pipe.txt | 68315
 22475    | tickit/allusers_pipe.txt  | 68061
   547    | date2008_pipe.txt         |  2572
 22487    | tickit/listings_pipe.txt  | 68072
  7531    | venue_pipe.txt            | 23390
  7583    | listings_pipe.txt         | 23445
(25 rows)
```

# SYS_LOAD_ERROR_DETAIL

Utilice SYS_LOAD_ERROR_DETAIL para visualizar los detalles de los errores del comando COPY. Cada una de las filas representa un comando COPY. Contiene tanto los comandos COPY en ejecución como los que han finalizado.

SYS_LOAD_ERROR_DETAIL es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | El identificador del usuario que envió la copia. |
| query_id | bigint | El identificador de consulta de la copia. |
| transaction_id | bigint | Identificador de la transacción. |
| session_id | entero | El identificador de proceso del proceso que ejecuta la copia. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| database_name | character(64) | El nombre de la base de datos a la que estaba conectado el usuario cuando se emitió la copia. |
| table_id | entero | El identificador de tabla. |
| start_time | marca de tiempo | La hora (UTC) en que comenzó la copia. |
| file_name | character (256) | La ruta completa hacia el archivo de entrada que hay que cargar. |
| line_number | bigint | El número de línea en el archivo de carga que tiene el error. Cuando se carga un archivo JSON, el número de línea de la última línea del objeto JSON que tiene el error. |
| column_name | character (127) | El campo que tiene el error. |
| column_type | character (10) | El tipo de datos del campo que tiene el error. |
| column_length | character (10) | La longitud de la columna, si corresponde. Este campo se rellena cuando el tipo de datos tiene una longitud limitada. Por ejemplo, para una columna con un tipo de datos “character (3)”, esta columna tiene el valor “3”. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| position | entero | La posición del error en el campo. |
| error_code | entero | Código de error. |
| error_message | character (512) | La explicación del error. |

## Consultas de ejemplo

La siguiente consulta muestra los detalles del error de carga del comando copy de una consulta específica.

```
SELECT query_id,
       table_id,
       start_time,
       trim(file_name) AS file_name,
       trim(column_name) AS column_name,
       trim(column_type) AS column_type,
       trim(error_message) AS error_message
FROM sys_load_error_detail
WHERE query_id = 762949
ORDER BY start_time
LIMIT 10;
```

Resultados de ejemplo.

```
 query_id | table_id |         start_time         |               file_name
      | column_name | column_type |                 error_message
----------+----------+----------------------------
+----------------------------------------+------------+------------
+-----------------------------------------------
   762949 |   137885 | 2022-02-15 22:14:46.759151 | s3://load-test/copyfail/
wrong_format_000 | id          | int4        | Invalid digit, Value 'a', Pos 0, Type:
 Integer
   762949 |   137885 | 2022-02-15 22:14:46.759151 | s3://load-test/copyfail/
wrong_format_001 | id          | int4        | Invalid digit, Value 'a', Pos 0, Type:
 Integer
```

# SYS_LOAD_HISTORY

Utilice SYS_LOAD_HISTORY para visualizar los detalles de los comandos COPY. Cada una de las filas representa un comando COPY con estadísticas acumuladas para algunos de los campos. Contiene tanto los comandos COPY en ejecución como los que han finalizado.

SYS_LOAD_HISTORY es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | El identificador del usuario que envió la copia. |
| query_id | bigint | El identificador de consulta de la copia. |
| transaction_id | bigint | Identificador de la transacción. |
| session_id | entero | El identificador de proceso del proceso que ejecuta la copia. |
| database_name | texto | El nombre de la base de datos a la que estaba conectado el usuario cuando se emitió la operación. |
| status | texto | El estado de la copia. Los valores válidos son `running`, `completed` , `aborted`. |
| table_name | texto | El nombre de la tabla en la que se copia. |
| start_time | marca de tiempo | La hora de inicio de la copia. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| end_time | marca de tiempo | La hora en que se completó la copia. |
| duración | bigint | La cantidad de tiempo (microsegundos) empleado en el comando COPY. |
| data_source | texto | La ubicación de Amazon S3 de los archivos de entrada que hay que copiar. |
| file_format | texto | El formato del archivo de origen. Los formatos incluyen csv, txt, json, avro, orc, o parquet. |
| loaded_rows | bigint | La cantidad de filas copiadas en la tabla. |
| loaded_bytes | bigint | La cantidad de bytes copiados en la tabla. |
| source_file_count | entero | La cantidad de archivos contabilizados en los archivos de origen. |
| source_file_bytes | bigint | La cantidad de bytes en los archivos de origen. |
| file_count_scanned | entero | La cantidad de archivos analizados de Amazon S3. |
| file_bytes_scanned | bigint | La cantidad de bytes analizados desde el archivo de Amazon S3. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| error_count | bigint | La cantidad de errores contabilizados. |
| copy_job_id | bigint | Identificador del trabajo de copia. Un 0 indica que no hay ningún identificador de trabajo. |

## Consultas de ejemplo

La siguiente consulta muestra las filas, los bytes, las tablas y el origen de datos cargados de comandos de copia específicos.

```
SELECT query_id,
       table_name,
       data_source,
       loaded_rows,
       loaded_bytes
FROM sys_load_history
WHERE query_id IN (6389,490791,441663,74374,72297)
ORDER BY query_id,
         data_source DESC;
```

Resultados de ejemplo.

```
 query_id |    table_name    |                                  data_source
              | loaded_rows | loaded_bytes
----------+------------------
+-------------------------------------------------------------------------+-------------
+---------------
     6389 | store_returns    | s3://load-test/data-sources/tpcds/2.8.0/textfile/1T/
store_returns/    |   287999764 | 1196240296158
    72297 | web_site         | s3://load-test/data-sources/tpcds/2.8.0/textfile/1T/
web_site/         |          54 |         43808
    74374 | ship_mode        | s3://load-test/data-sources/tpcds/2.8.0/textfile/1T/
ship_mode/        |          20 |          1320
   441663 | income_band      | s3://load-test/data-sources/tpcds/2.8.0/textfile/1T/
income_band/      |          20 |          2152
```

```
   490791 | customer_address | s3://load-test/data-sources/tpcds/2.8.0/textfile/1T/
customer_address/ |     6000000 |     722924305
```

La siguiente consulta muestra las filas, los bytes, las tablas y el origen de datos cargados de comandos copy.

```
SELECT query_id,
       table_name,
       data_source,
       loaded_rows,
       loaded_bytes
FROM sys_load_history
ORDER BY query_id DESC
LIMIT 10;
```

Resultados de ejemplo.

```
 query_id |       table_name       |                                data_source
                          | loaded_rows |  loaded_bytes
----------+------------------------
+-------------------------------------------------------------------------------
+-------------+-----------------
   491058 | web_site               | s3://load-test/data-sources/tpcds/2.8.0/
textfile/1T/web_site/               |          54 |           43808
   490947 | web_sales              | s3://load-test/data-sources/tpcds/2.8.0/
textfile/1T/web_sales/              |   720000376 |  22971988122819
   490923 | web_returns            | s3://load-test/data-sources/tpcds/2.8.0/
textfile/1T/web_returns/            |    71997522 |     96597496325
   490918 | web_page               | s3://load-test/data-sources/tpcds/2.8.0/
textfile/1T/web_page/               |        3000 |            1320
   490907 | warehouse              | s3://load-test/data-sources/tpcds/2.8.0/
textfile/1T/warehouse/              |          20 |            1320
   490902 | time_dim               | s3://load-test/data-sources/tpcds/2.8.0/
textfile/1T/time_dim/               |       86400 |            1320
   490876 | store_sales            | s3://load-test/data-sources/tpcds/2.8.0/
textfile/1T/store_sales/            |  2879987999 | 151666241887933
   490870 | store_returns          | s3://load-test/data-sources/tpcds/2.8.0/
textfile/1T/store_returns/          |   287999764 |   1196405607941
   490865 | store                  | s3://load-test/data-sources/tpcds/2.8.0/
textfile/1T/store/                  |        1002 |          365507
```

La siguiente consulta muestra las filas y los bytes cargados del comando copy.

```
SELECT date_trunc('day',start_time) AS exec_day,
       SUM(loaded_rows) AS loaded_rows,
       SUM(loaded_bytes) AS loaded_bytes
FROM sys_load_history
GROUP BY exec_day
ORDER BY exec_day DESC;
```

Resultados de ejemplo.

```
      exec_day       | loaded_rows |   loaded_bytes
---------------------+-------------+------------------
 2022-01-20 00:00:00 |  6347386005 |  258329473070606
 2022-01-19 00:00:00 | 19042158015 |  775198502204572
 2022-01-18 00:00:00 | 38084316030 | 1550294469446883
 2022-01-17 00:00:00 | 25389544020 | 1033271084791724
 2022-01-16 00:00:00 | 19042158015 |  775222736252792
 2022-01-15 00:00:00 | 19834245387 |  798122849155598
 2022-01-14 00:00:00 | 75376544688 | 3077040926571384
```

# SYS_MV_REFRESH_HISTORY

Los resultados incluyen información sobre el historial de actualizaciones de todas las vistas materializadas. Los resultados incluyen el tipo de actualización, como manual o automática, y el estado de la actualización más reciente.

SYS_MV_REFRESH_HISTORY está visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | El identificador del usuario que envió la actualización. |
| session_id | entero | El identificador del proceso que ejecuta la actualización de la vista materializada. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| transaction_id | bigint | Identificador de la transacción. |
| database_name | char(128) | Base de datos que contiene la vista materializada. |
| schema_name | char(128) | Esquema de la vista materiali zada. |
| mv_id | bigint | ID de objeto de la vista materializada. |
| mv_name | char(128) | El nombre de la vista materiali zada. |
| refresh_type | char(32) | El tipo de actualización, por ejemplo, manual o automática. |
| status | texto | El estado de la actualización. Para obtener información detallada sobre los estados, consulte la columna de estado de SVL_MV_REFRESH_STATUS. |
| start_time | marca de tiempo | La hora de inicio de la actualización. |
| end_time | marca de tiempo | La hora de finalización de la actualización. |
| duración | bigint | La cantidad de tiempo en microsegundos que se tardó en actualizar la vista materiali zada. |

## Consultas de ejemplo

La siguiente consulta muestra el historial de actualizaciones de las vistas materializadas.

```
SELECT user_id,
     session_id,
     transaction_id,
     database_name,
     schema_name,
     mv_id,
     mv_name,
     refresh_type,
     status,
     start_time,
     end_time,
     duration
     from sys_mv_refresh_history;
```

La consulta devuelve el siguiente ejemplo de salida:

```
 user_id | session_id | transaction_id | database_name | schema_name                |
 mv_id  |  mv_name           |  refresh_type  |  status
         |  start_time                |  end_time                  |  duration
---------+------------+----------------+---------------
+----------------------------+--------+--------------------+----------------
+-------------------------------------------------+----------------------------
+----------------------------+-----------
       1 | 1073815659 |          15066 | dev           | test_stl_mv_refresh_schema |
 203762 | mv_incremental     | Manual         | MV was already updated
         | 2023-10-26 15:59:20.952179 | 2023-10-26 15:59:20.952866 |      687
       1 | 1073815659 |          15068 | dev           | test_stl_mv_refresh_schema |
 203771 | mv_nonincremental  | Manual         | MV was already updated
         | 2023-10-26 15:59:21.008049 | 2023-10-26 15:59:21.008658 |      609
       1 | 1073815659 |          15070 | dev           | test_stl_mv_refresh_schema |
 203779 | mv_refresh_error   | Manual         | MV was already updated
         | 2023-10-26 15:59:21.064252 | 2023-10-26 15:59:21.064885 |      633
       1 | 1073815659 |          15074 | dev           | test_stl_mv_refresh_schema
 | 203762 | mv_incremental     | Manual         | Refresh successfully updated MV
 incrementally   | 2023-10-26 15:59:29.693329 | 2023-10-26 15:59:43.482842 | 13789513
       1 | 1073815659 |          15076 | dev           | test_stl_mv_refresh_schema |
 203771 | mv_nonincremental  | Manual         | Refresh successfully recomputed MV from
 scratch | 2023-10-26 15:59:43.550184 | 2023-10-26 15:59:47.880833 |  4330649
```

```
       1 | 1073815659 |          15078 | dev           | test_stl_mv_refresh_schema |
 203779 | mv_refresh_error   | Manual         | Refresh failed due to an internal error
        | 2023-10-26 15:59:47.949052 | 2023-10-26 15:59:52.494681 |  4545629
(6 rows)
```

# SYS_MV_STATE

Los resultados incluyen información sobre el estado de todas las vistas materializadas. Incluye información de la tabla base, propiedades del esquema e información sobre eventos recientes, como la eliminación de una columna.

SYS_MV_STATE es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | bigint | El ID del usuario que creó el evento. |
| transaction_id | bigint | ID de transacción del evento. |
| database_name | char(128) | Base de datos que contiene la vista materializada. |
| event_desc | char(500) | El evento que solicitó el cambio de estado. Entre los valores de ejemplo se incluyen:<br>• Se ha cambiado el tipo de columna<br>• Se ha eliminado la columna<br>• Se ha cambiado el nombre de la columna<br>• Se ha cambiado el nombre del esquema |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | • Conversión de tabla pequeña<br>• TRUNCATE<br>• Vacuum<br><br>Tenga en cuenta que existen otros valores posibles para esta columna. |
| start_time | marca de tiempo | La hora de inicio del evento. |
| base_table_database_name | char(128) | Nombre de la base de datos de la tabla base. |
| base_table_schema | char(128) | Esquema de la tabla base. |
| base_table_name | char(128) | Nombre de la tabla base. |
| mv_schema | char(128) | Esquema de la vista materiali zada. |
| mv_name | char(128) | El nombre de la vista materiali zada. |
| estado | character (32) | El estado cambiado de la vista materializada, que es el siguiente:<br>• Volver a calcular<br>• No se puede actualizar |

## Consultas de ejemplo

En la siguiente consulta se muestra el estado de la vista materializada.

```
select * from sys_mv_state;
```

La consulta devuelve el siguiente ejemplo de salida:

```
 user_id | transaction_id | database_name | event_desc                   | start_time
              | base_table_database_name | base_table_schema | base_table_name     |
 mv_schema  | mv_name                    | state
---------+----------------+---------------+------------------------------
+--------------------------+-------------------------+------------------
+--------------------+------------+---------------------------+-------------
 106     | 12720          | tickit_db     | TRUNCATE                    | 2023-07-26
 14:59:12.788268 | tickit_db                | mv_schema         | test_table_95d6d861 |
 mv_schema   | materialized_view_a1f3f862 | Recompute
 106     | 12724          | tickit_db     | ALTER TABLE ALTER DISTSTYLE | 2023-07-26
 14:59:51.409014 | tickit_db                | mv_schema         | test_table_58102435 |
 mv_schema   | materialized_view_ca746631 | Recompute
 106     | 12720          | tickit_db     | Column was renamed          | 2023-07-26
 14:59:12.822928 | tickit_db                | mv_schema         | test_table_95d6d861 |
 mv_schema   | materialized_view_5750a8d4 | Unrefreshable
 106     | 12727          | tickit_db     | Table was renamed           | 2023-07-26
 15:00:08.051244 | tickit_db                | mv_schema         | test_table_95d6d861 |
 mv_schema   | materialized_view_5750a8d4 | Unrefreshable
 106     | 12720          | tickit_db     | Column was renamed          | 2023-07-26
 14:59:12.857755 | tickit_db                | mv_schema         | test_table_95d6d861 |
 mv_schema   | materialized_view_5750a8d4 | Unrefreshable
 106     | 12727          | tickit_db     | Table was renamed           | 2023-07-26
 15:00:08.051358 | tickit_db                | mv_schema         | test_table_95d6d861 |
 mv_schema   | materialized_view_5ef0d754 | Unrefreshable
 106     | 12720          | tickit_db     | TRUNCATE                    | 2023-07-26
 14:59:12.788159 | tickit_db                | mv_schema         | test_table_95d6d861 |
 mv_schema   | materialized_view_5750a8d4 | Recompute
 106     | 12720          | tickit_db     | Column was renamed          | 2023-07-26
 14:59:12.857799 | tickit_db                | mv_schema         | test_table_95d6d861 |
 mv_schema   | materialized_view_a1f3f862 | Unrefreshable
 106     | 12720          | tickit_db     | TRUNCATE                    | 2023-07-26
 14:59:12.788327 | tickit_db                | mv_schema         | test_table_95d6d861 |
 mv_schema   | materialized_view_5ef0d754 | Recompute
 106     | 12727          | tickit_db     | ALTER TABLE ALTER SORTKEY   | 2023-07-26
 15:00:08.006235 | tickit_db                | mv_schema         | test_table_58102435 |
 mv_schema   | materialized_view_ca746631 | Recompute
 106     | 12720          | tickit_db     | Column was renamed          | 2023-07-26
 14:59:12.82297  | tickit_db                | mv_schema         | test_table_95d6d861 |
 mv_schema   | materialized_view_a1f3f862 | Unrefreshable
```

```
106     | 12727          | tickit_db     | Table was renamed           | 2023-07-26
15:00:08.051321 | tickit_db                | mv_schema         | test_table_95d6d861 |
mv_schema   | materialized_view_a1f3f862 | Unrefreshable
```

# SYS_PROCEDURE_CALL

Utilice la vista SYS_PROCEDURE_CALL para obtener información sobre las llamadas a procedimientos almacenados, incluida la hora de inicio, la hora de finalización, el estado de una llamada a un procedimiento almacenado y la jerarquía de llamadas para las llamadas a procedimientos almacenados anidados. Cada llamada de procedimiento almacenado recibe un ID de consulta.

SYS_STORED_PROC_CALL es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| session_user_id | entero | Identificador del usuario que ha creado la sesión y es el invocador de la llamada al procedimiento almacenado de nivel superior. |
| security_user_id | entero | Identificador del usuario cuyos privilegios se utilizaron para ejecutar la instrucción en el procedimiento almacenado. Si el procedimiento almacenad o se hubiera llamado desde el procedimiento almacenado DEFINER exterior, este será el user_id del propietario del procedimiento almacenado exterior. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| query_id | entero | El identificador de consulta de la llamada al procedimiento almacenado. |
| query_text | char(4000) | El texto de la consulta de llamada al procedimiento almacenado. |
| start_time | marca de tiempo | La hora en UTC en la que comenzó a ejecutarse la consulta. Por ejemplo, la marca temporal utiliza seis dígitos de precisión para los segundos fraccionarios. 2009-06-12 11:29:19.131358. |
| end_time | marca de tiempo | La hora en UTC a la que terminó de ejecutarse la consulta. La marca temporal utiliza seis dígitos de precisión para los segundos fracciona rios, por ejemplo: 2009-06-12 11:29:19.131358. |
| status | char(10) | El estado de la llamada al procedimiento almacenado. Cuando el sistema detiene el procedimiento almacenad o o el usuario lo cancela, se anula el valor. Si la llamada al procedimiento almacenado se ejecuta hasta completarse, el valor es correcto. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| caller_procedure_query_id | entero | Si la llamada al procedimiento almacenado la invocó otra llamada a procedimiento almacenado, esta columna contiene el ID de consulta de la llamada externa. De lo contrario, el campo es NULL. |

### Consultas de ejemplo

La siguiente consulta devuelve una jerarquía de llamadas a procedimientos almacenados anidados.

```
select query_id, datediff(seconds, start_time, end_time) as elapsed_time, status,
 trim(query_text) as call, caller_procedure_query_id from sys_procedure_call;
```

Resultados de ejemplo.

```
 query_id | elapsed_time | status  |                      call                      |
 caller_procedure_query_id
----------+--------------+---------+------------------------------------------------
+---------------------------
     3087 |           18 | success | CALL proc_bd906c98c45443ffa165e9552056902d(1)    |
          3085
     3085 |           18 | success | CALL proc_bd906c98c45443ffa165e9552056902d_2(1); |

(2 rows)
```

## SYS_PROCEDURE_MESSAGES

SYS_STORED_PROC_MESSAGES es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| transaction_id | bigint | Identificador de la transacción. |
| query_id | entero | El identificador de consulta de la llamada al procedimiento almacenado. |
| record_time | marca de tiempo | La hora en UTC en la que se generó el mensaje. |
| log_level | char(10) | El nivel de registro del mensaje generado. Los valores posibles son: LOG, INFO, NOTICE, WARNING y EXCEPTION. |
| message | char (1024) | El texto del mensaje generado. |
| line_number | entero | El número de línea del mensaje generado. |

## Consultas de ejemplo

La siguiente consulta muestra un ejemplo de salida de SYS_PROCEDURE_MESSAGES.

```
select transaction_id, query_id, record_time, log_level, trim(message), line_number
 from sys_procedure_messages;
```

```
transaction_id | query_id |        record_time         | log_level |           btrim
        | line_number
---------------+----------+----------------------------+-----------
+---------------------------+-------------
     25267     |   80562  | 2023-07-17 14:38:31.910136 |   NOTICE  |
 test_notice_msg_b9f1e749  |     8
```

```
     25267     |   80562  | 2023-07-17 14:38:31.910002 |    LOG    |
 test_log_msg_833c7420    |     6
     25267     |   80562  | 2023-07-17 14:38:31.910111 |    INFO   |
 test_info_msg_651373d9   |     7
     25267     |   80562  | 2023-07-17 14:38:31.910154 |   WARNING |
 test_warning_msg_831c5747 |     9
(4 rows)
```

## SYS_QUERY_DETAIL

Utilice SYS_QUERY_DETAIL para visualizar los detalles de las consultas en los distintos niveles de las métricas, donde cada fila representa los detalles de una consulta de WLM en un nivel de la métrica determinado. Esta vista contiene muchos tipos de consultas como DDL, DML y comandos de utilidad (por ejemplo, copiar y descargar). Según el tipo de consulta, algunas columnas pueden resultar irrelevantes. Por ejemplo, external_scanned_bytes no es relevante para las tablas internas.

SYS_QUERY_DETAIL es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> Para verificar si una transacción que contiene la consulta ejecutada se ha confirmado correctamente, debe realizar una operación de unión entre las tablas del sistema y la tabla `sys_transaction_history`. Por ejemplo:
>
> ```
> SELECT
>     th.transaction_id,
>     qd.query_id,
>     th.status AS transaction_status
> FROM
>     sys_query_detail qd
> LEFT JOIN sys_query_history qh ON qd.query_id = qh.query_id
> LEFT JOIN sys_transaction_history th on qh.transaction_id = th.transaction_id;
> ```

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | El identificador del usuario que envió la consulta. |
| query_id | bigint | Identificador de la consulta. |
| child_query_sequence | entero | La secuencia de la consulta reescrita del usuario, a partir de 1. |
| stream_id | entero | El identificador del flujo de la secuencia de la consulta. |
| segment_id | entero | El identificador del segmento de ejecución de la consulta. |
| step_id | entero | El identificador del paso en un segmento. |
| step_name | texto | El nombre del paso en un segmento. Los posibles valores son `aggregate` , `broadcast` , `delete`, `distribute` , `hash`, `hashjoin`, `insert`, `limit`, `merge`, `nestloop`, `parse`, `return`, `save`, `scan`, `sort`, `sortlimit` , `unique` y `window`. |
| table_id | entero | El identificador de la tabla para realizar análisis permanentes de la tabla. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| table_name | character (136) | Nombre de la tabla del paso que está en curso. |
| is_rrscan | character | El valor que indica si un paso es un paso de análisis. El valor true (t) (verdadero) indica que se ha empleado un análisis de rango restringido. |
| start_time | marca de tiempo | La hora de inicio del paso de la consulta. Este campo se registra en el segmento, independientemente del valor de la columna `metrics_level` . |
| end_time | marca de tiempo | La hora en que se completó el paso de la consulta. Este campo se registra en el segmento, independientemente del valor de la columna `metrics_level` . |
| duración | bigint | La cantidad de tiempo (microsegundos) empleado en el paso. Este campo se registra en el segmento, independientemente del valor de la columna `metrics_level` . |
| alerta | texto | La descripción del evento de alerta. |
| input_bytes | bigint | Los bytes de entrada del paso actual. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| input_rows | bigint | Las filas de entrada del paso actual. |
| output_bytes | bigint | Los bytes de salida del paso actual. |
| output_rows | bigint | Las filas de salida del paso actual. |
| blocks_read | bigint | El número de bloque que ha leído el paso. |
| blocks_write | bigint | El número de bloque que ha escrito el paso. |
| local_read_IO | bigint | El número de bloques leídos de la memoria caché del disco local. |
| remote_read_IO | bigint | El número de bloques leídos de forma remota. |
| origen | texto | El tipo de objeto de base de datos que se escaneó. Esta columna solo tiene un valor cuando el valor step_name de la fila es `scan`. |
| data_skewness | entero | La asimetría de la distribuc ión de las filas de salida entre todos los pasos. Es un número comprendido entre el 0 % y el 100 %. Cuanto mayor sea el número, más desequili brada está la distribución. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| time_skewness | entero | La asimetría de la distribución del tiempo de ejecución entre todos los pasos. Es un número comprendido entre el 0 % y el 100 %. Cuanto mayor sea el número, más desequilibrada está la distribución. |
| is_active | character | Se trata del estado de la consulta en el nivel de paso. Los valores posibles son «t», que indica que el paso se está ejecutando activamente, o «f», que indica que el paso ha terminado de ejecutarse. |
| spilled_block_local_disk | bigint | La cantidad de bloques vertidos en el disco local. |
| spilled_block_remote_disk | bigint | La cantidad de bloques vertidos en Amazon Simple Storage Service. |
| step_attribute | character(64) | Contiene información sobre el paso asociado. Valores posibles para los pasos de escaneo: `multi-dimensional` . |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| metrics_level | character(64) | Nivel de la métrica de la consulta. Los valores posibles son los siguientes:<br>• child query<br>• secuencia<br>• segment<br>• paso<br>• |
| plan_parent_id | entero | Identificador del nodo principal del nodo del plan. Un nodo principal puede tener varios nodos secundarios. Por ejemplo, una unión de combinación es el nodo principal de los análisis de las tablas unidas. |
| plan_node_id | entero | Identificador de un nodo del plan que se asigna a uno o más pasos de la consulta. |

## Notas de uso

SYS_QUERY_DETAIL puede contener métricas en el paso, la secuencia, el segmento y la consulta secundaria. Además de hacer referencia a la columna metrics_level, puede ver qué nivel de métrica muestra una fila determinada haciendo referencia a los campos step_id, segment_id y stream_id según la siguiente tabla.

| Nivel de métricas | stream_id value | segment_id value | step_id value |
|---|---|---|---|
| child query | -1 | -1 | -1 |

| Nivel de métricas | stream_id value | segment_id value | step_id value |
|---|---|---|---|
| secuencia | Valor de paso válido | -1 | -1 |
| segment | Valor de paso válido | Valor de paso válido | -1 |
| paso | Valor de paso válido | Valor de paso válido | Valor de paso válido |

## Consultas de ejemplo

En el siguiente ejemplo, se devuelve el resultado de SYS_QUERY_DETAIL.

La siguiente consulta muestra los detalles de metadatos de la consulta con respecto al paso, incluidos el nombre del paso, input_bytes, output_bytes, input_rows y output_rows.

```
SELECT query_id,
       child_query_sequence,
       stream_id,
       segment_id,
       step_id,
       trim(step_name) AS step_name,
       duration,
       input_bytes,
       output_bytes,
       input_rows,
       output_rows
FROM sys_query_detail
WHERE query_id IN (193929)
ORDER BY query_id,
         stream_id,
         segment_id,
         step_id DESC;
```

Resultados de ejemplo.

```
 query_id | child_query_sequence | stream_id | segment_id | step_id | step_name  |
 duration     | input_bytes | output_bytes | input_rows | output_rows
----------+----------------------+-----------+------------+---------+------------
+-----------------+-------------+--------------+------------+-------------
   193929 |                    2 |         0 |          0 |       3 | hash       |
      37144 |           0 |      9350272 |          0 |      292196
```

```
193929 |                   5 |         0 |          0 |       3 | hash       |
    9492 |           0 |        23360 |          0 |        1460
193929 |                   1 |         0 |          0 |       3 | hash       |
   46809 |           0 |      9350272 |          0 |      292196
193929 |                   4 |         0 |          0 |       2 | return     |
    7685 |           0 |          896 |          0 |         112
193929 |                   1 |         0 |          0 |       2 | project    |
   46809 |           0 |            0 |          0 |      292196
193929 |                   2 |         0 |          0 |       2 | project    |
   37144 |           0 |            0 |          0 |      292196
193929 |                   5 |         0 |          0 |       2 | project    |
    9492 |           0 |            0 |          0 |        1460
193929 |                   3 |         0 |          0 |       2 | return     |
   11033 |           0 |        14336 |          0 |         112
193929 |                   2 |         0 |          0 |       1 | project    |
   37144 |           0 |            0 |          0 |      292196
193929 |                   1 |         0 |          0 |       1 | project    |
   46809 |           0 |            0 |          0 |      292196
193929 |                   5 |         0 |          0 |       1 | project    |
    9492 |           0 |            0 |          0 |        1460
193929 |                   3 |         0 |          0 |       1 | aggregate  |
   11033 |           0 |       201488 |          0 |          14
193929 |                   4 |         0 |          0 |       1 | aggregate  |
    7685 |           0 |        28784 |          0 |          14
193929 |                   5 |         0 |          0 |       0 | scan       |
    9492 |           0 |        23360 |     292196 |        1460
193929 |                   4 |         0 |          0 |       0 | scan       |
    7685 |           0 |         1344 |        112 |         112
193929 |                   2 |         0 |          0 |       0 | scan       |
   37144 |           0 |      7304900 |     292196 |      292196
193929 |                   3 |         0 |          0 |       0 | scan       |
   11033 |           0 |        13440 |        112 |         112
193929 |                   1 |         0 |          0 |       0 | scan       |
   46809 |           0 |      7304900 |     292196 |      292196
193929 |                   5 |         0 |          0 |      -1 |            |
    9492 |       12288 |            0 |          0 |          0
193929 |                   1 |         0 |          0 |      -1 |            |
   46809 |       16384 |            0 |          0 |          0
193929 |                   2 |         0 |          0 |      -1 |            |
   37144 |       16384 |            0 |          0 |          0
193929 |                   4 |         0 |          0 |      -1 |            |
    7685 |       28672 |            0 |          0 |          0
```

```
   193929 |                   3 |         0 |          0 |      -1 |           |
      11033 |      114688 |            0 |          0 |           0
```

Para ver las tablas de la base de datos ordenadas de las más utilizadas a las menos utilizadas, utilice el siguiente ejemplo. Reemplace *sample_data_dev* con su propia base de datos. Tenga en cuenta que esta consulta contará las consultas a partir del momento en que se cree el clúster, pero los datos de visualización del sistema no se guardarán cuando no haya espacio en el almacenamiento de datos.

```
SELECT table_name, COUNT (DISTINCT query_id)
FROM SYS_QUERY_DETAIL
WHERE table_name LIKE 'sample_data_dev%'
GROUP BY table_name
ORDER BY COUNT(*) DESC;

+---------------------------------+-------+
|           table_name            | count |
+---------------------------------+-------+
| sample_data_dev.tickit.venue    |     4 |
| sample_data_dev.myunload1.venue |     3 |
| sample_data_dev.tickit.listing  |     1 |
| sample_data_dev.tickit.category |     1 |
| sample_data_dev.tickit.users    |     1 |
| sample_data_dev.tickit.date     |     1 |
| sample_data_dev.tickit.sales    |     1 |
| sample_data_dev.tickit.event    |     1 |
+---------------------------------+-------+
```

En el siguiente ejemplo, se muestran los distintos niveles de métricas de una única consulta de WLM.

```
SELECT query_id, child_query_sequence, stream_id, segment_id, step_id, step_name,
 start_time, end_time, metrics_level
FROM sys_query_detail
WHERE query_id = 1553 AND step_id = -1
ORDER BY stream_id, segment_id, step_id;

 query_id | child_query_sequence | stream_id | segment_id | step_id | step_name |
   start_time         |          end_time          | metrics_level
----------+----------------------+-----------+------------+---------+-----------
+----------------------------+----------------------------+---------------
     1553 |                    1 |        -1 |         -1 |      -1 |           |
 2024-10-17 02:28:49.814721 | 2024-10-17 02:28:49.847838 | child query
```

```
    1553 |                    1 |         0 |         -1 |      -1 |           |
 2024-10-17 02:28:49.814721 | 2024-10-17 02:28:49.835609 | stream
    1553 |                    1 |         0 |          0 |      -1 |           |
 2024-10-17 02:28:49.824677 | 2024-10-17 02:28:49.830372 | segment
    1553 |                    1 |         1 |         -1 |      -1 |           |
 2024-10-17 02:28:49.835624 | 2024-10-17 02:28:49.845773 | stream
    1553 |                    1 |         1 |          1 |      -1 |           |
 2024-10-17 02:28:49.84088  | 2024-10-17 02:28:49.842388 | segment
    1553 |                    1 |         1 |          2 |      -1 |           |
 2024-10-17 02:28:49.835926 | 2024-10-17 02:28:49.844396 | segment
    1553 |                    1 |         2 |         -1 |      -1 |           |
 2024-10-17 02:28:49.846949 | 2024-10-17 02:28:49.847838 | stream
    1553 |                    1 |         2 |          3 |      -1 |           |
 2024-10-17 02:28:49.847013 | 2024-10-17 02:28:49.847485 | segment
(8 rows)
```

# SYS_QUERY_EXPLAIN

Muestra el plan EXPLAIN de una consulta que se ha enviado para ejecutarse.

SYS_QUERY_EXPLAIN es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | El identificador del usuario que envió la consulta. |
| query_id | bigint | Identificador de la consulta. La información detallada de las consultas se almacena en SYS_QUERY_HISTORY. |
| child_query_sequence | entero | La secuencia de la consulta reescrita del usuario, a partir de 1. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| plan_node_id | entero | Identificador de un nodo del plan que se asigna a uno o más pasos de la consulta. |
| plan_parent_id | entero | Identificador del nodo principal del nodo del plan. Un nodo principal puede tener varios nodos secundarios. Por ejemplo, una unión de combinación es el nodo principal de los análisis de las tablas unidas. |
| plan_node | character (400) | El texto del nodo extraído de la salida de EXPLAIN. Los nodos del plan que hacen referencia a la ejecución en nodos de computación llevan el prefijo XN en la salida de EXPLAIN. |
| node_info | character (400) | Información del calificador y del filtro para el nodo del plan. Por ejemplo, las condicion es de combinación y las restricciones de la cláusula WHERE están incluidas en esta columna. |

## Consultas de ejemplo

El siguiente ejemplo es el plan EXPLAIN de una sola consulta.

```
SELECT * FROM sys_query_explain WHERE query_id = 612635 ORDER_BY plan_node_id;
```

```
 userid | query_id | child_query_sequence | plan_node_id | plan_parent_id |

          plan_node

                                          |

                                                                  plan_info


--------+----------+----------------------+--------------+----------------
+-----------------------------------------------------------------------------------------------------
+-----------------------------------------------------------------------------------------------------
    100 |   612635 |                    1 |            1 |              0 | XN Limit
 (cost=3604047533041.00..3604047533041.25 rows=100 width=20)

                                            |




    100 |   612635 |                    1 |            2 |              1 |   ->  XN
 Merge  (cost=3604047533041.00..3604047533148.02 rows=42809 width=20)

                                             | Merge Key: sum(b.totalprice)




    100 |   612635 |                    1 |            3 |              2 |         ->
 XN Network  (cost=3604047533041.00..3604047533148.02 rows=42809 width=20)

                                          |   Send to leader
```

```
  100 |   612635 |                    1 |           4 |             3 |
 ->  XN Sort  (cost=3604047533041.00..3604047533148.02 rows=42809 width=20)

                                        | Sort Key: sum(b.totalprice)

  100 |   612635 |                    1 |           5 |             4 |
          ->  XN HashAggregate  (cost=2604047529640.76..2604047529747.78 rows=42809
width=20)

                                              |

  100 |   612635 |                    1 |           6 |             5 |
                ->  XN Hash Join DS_DIST_NONE  (cost=15104956.16..2602364653507.34
rows=336575226684 width=20)

                                                     | Hash Cond: (("outer".listid =
"inner".listid) AND ("outer".sellerid = "inner".sellerid))

  100 |   612635 |                    1 |           7 |             6 |
                          ->  XN Seq Scan on listing b  (cost=0.00..7884677.12
rows=788467712 width=24)

                                                       |
```

```
   100 |   612635 |                    1 |           8 |             6 |
                  ->  XN Hash  (cost=7063797.76..7063797.76 rows=706379776 width=8)

                                        |

   100 |   612635 |                    1 |           9 |             8 |
                             ->  XN Seq Scan on sales a  (cost=0.00..7063797.76
 rows=706379776 width=8)

                                                  |

(9 rows)
```

## SYS_QUERY_HISTORY

Utilice SYS_QUERY_HISTORY para visualizar los detalles de las consultas de los usuarios. Cada una de las filas representa una consulta del usuario con estadísticas acumuladas para algunos de los campos. Esta vista contiene muchos tipos de consultas, como por ejemplo lenguaje de definición de datos (DDL), lenguaje de manipulación de datos (DML), copia, descarga y Amazon Redshift Spectrum. Contiene tanto las consultas en curso como las que han finalizado.

SYS_QUERY_HISTORY es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> Para verificar si una transacción que contiene la consulta ejecutada se ha confirmado correctamente, debe realizar una operación de unión entre las tablas del sistema y la tabla `sys_transaction_history`. Por ejemplo:

```
SELECT
    qh.transaction_id,
    qh.query_id,
    qh.status AS query_status,
    qh.query_type,
    TRIM(qh.query_text) AS query_text,
    th.status AS transaction_status
FROM
    sys_query_history qh
LEFT JOIN
    sys_transaction_history th ON qh.transaction_id = th.transaction_id;
```

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | El identificador del usuario que envió la consulta. |
| query_id | bigint | Identificador de la consulta. |
| query_label | character(320) | El nombre abreviado de la consulta. |
| transaction_id | bigint | Identificador de la transacción. |
| session_id | entero | El identificador del proceso que ejecuta la consulta. |
| database_name | character (128) | El nombre de la base de datos al que estaba conectado el |

| Nombre de la columna | Tipo de datos: | Descripción |
| --- | --- | --- |
| | | usuario cuando se emitió la consulta. |
| query_type | character (32) | El tipo de consulta, como, SELECT, INSERT, UPDATE, UNLOAD, COPY, COMMAND, DDL, UTILITY, CTAS y OTHER. |
| status | character (10) | El estado de la consulta. Valores válidos: planificación, en cola, en ejecución, en retorno, erróneo, cancelado y correcto. |
| result_cache_hit | Booleano | Indica si la consulta coincide con la memoria caché de resultados. |
| start_time | marca de tiempo | La hora de inicio de la consulta. |
| end_time | marca de tiempo | La hora en que se completó la consulta. |
| elapsed_time | bigint | El tiempo total (microseg undos) empleado en la consulta. |
| queue_time | bigint | El tiempo total (microseg undos) empleado en la cola de consulta de la clase de servicio. |
| execution_time | bigint | El tiempo total (microseg undos) que se ejecuta en la clase de servicio. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| error_message | character (512) | El motivo por el cual se produjo un error en la consulta. |
| returned_rows | bigint | La cantidad de filas devueltas al cliente. |
| returned_bytes | bigint | La cantidad de bytes devueltos al cliente. |
| query_text | character (4000) | La cadena de consulta. Es posible que esta cadena quede truncada. |
| redshift_version | character (256) | La versión de Amazon Redshift cuando se ejecutó la consulta. |
| usage_limit | character(150) | Lista de los ID de límite de uso alcanzados por la consulta. |
| compute_type | varchar (32) | Indica si la consulta se ejecuta en el clúster principal o en un clúster de escalado de simultaneidad. Los valores posibles son `primary` (la consulta se ejecuta en el clúster principal), `secondary` (la consulta se ejecuta en el clúster secundario) o `primary-scale` (la consulta se ejecuta en el clúster de simultaneidad). Esto solo se aplica al clúster aprovisionado. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| compile_time | bigint | El tiempo total (microsegundos) empleado en la compilación de la consulta. |
| planning_time | bigint | El tiempo total (microsegundos) empleado en la planificación de la consulta. |
| lock_wait_time | bigint | El tiempo total (microsegundos) empleado en la espera del bloqueo de relación. |
| service_class_id | entero | ID de la clase de servicio. Para obtener una lista de ID de clase de servicio, vaya a ID de la clase de servicio WLM.<br><br>Esta columna solo se utiliza para consultas ejecutadas en clústeres aprovisionados. Para las consultas que se ejecutan en Redshift sin servidor, esta columna contiene -1. |
| service_class_name | character(64) | Es el nombre de la clase de servicio.<br><br>Esta columna solo se utiliza para consultas ejecutadas en clústeres aprovisionados. Para las consultas que se ejecutan en Amazon Redshift sin servidor, esta columna está vacía. |

| Nombre de la columna | Tipo de datos: | Descripción |
| --- | --- | --- |
| query_priority | character(20) | La prioridad de la cola en la que se ejecutó la consulta. Los valores posibles son los siguientes:<br>• NULL<br>• lowest<br>• low<br>• normal<br>• high<br>• highest<br><br>NULL significa que la consulta no admite la prioridad de consulta.<br><br>Esta columna solo se utiliza para consultas ejecutadas en clústeres aprovisionados. Para las consultas que se ejecutan en Redshift sin servidor, esta columna está vacía. |

| Nombre de la columna | Tipo de datos: | Descripción |
| --- | --- | --- |
| short_query_accelerated | character (10) | Indica si la consulta se ha acelerado mediante la aceleración de consultas cortas (SQA). Los valores posibles son los siguientes:<br>• true<br>• false<br>• NULL<br><br>Esta columna solo se utiliza para consultas ejecutadas en clústeres aprovisionados. Para las consultas que se ejecutan en Redshift sin servidor, esta columna está vacía. |
| user_query_hash | character(40) | El hash de consulta generado a partir de la consulta, incluidos los literales de consulta. Las consultas repetidas con el mismo texto de consulta tendrán los mismos valores de user_quer y_hash. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| generic_query_hash | character(40) | El hash de consulta generado a partir de la consulta, excluidos los literales de consulta. Las consultas repetidas con el mismo texto de consulta, pero diferentes literales de consulta, tendrán los mismos valores de generic_query_hash. |
| query_hash_version | entero | El número de versión del hash de consulta generado a partir de la consulta. |
| result_cache_query_id | entero | Si la consulta ha utilizado el almacenamiento en caché de los resultados, este valor de campo es el ID de la consulta de origen de los resultados almacenados en caché. Si no se utilizó el almacenamiento en caché, el valor de este campo es `0`. |
| nombre de usuario | character (128) | El nombre de usuario del usuario que envió la consulta. |

## Consultas de ejemplo

La siguiente consulta devuelve consultas en ejecución y en cola.

```
SELECT user_id,
       query_id,
       transaction_id,
       session_id,
       status,
```

```
       trim(database_name) AS database_name,
       start_time,
       end_time,
       result_cache_hit,
       elapsed_time,
       queue_time,
       execution_time
FROM sys_query_history
WHERE status IN ('running','queued')
ORDER BY start_time;
```

Resultados de ejemplo.

```
 user_id | query_id | transaction_id | session_id | status  | database_name |
 start_time         |          end_time          | result_cache_hit | elapsed_time |
 queue_time | execution_time
---------+----------+----------------+------------+---------+---------------
+---------------------------+---------------------------+-----------------
+--------------+------------+----------------
     101 |   760705 |         852337 | 1073832321 | running | tpcds_1t      |
 2022-02-15 19:03:19.67849 | 2022-02-15 19:03:19.739811 | f               |
 61321 |          0 |              0
```

La siguiente consulta devuelve la hora de inicio, la hora de finalización, el tiempo en cola, el tiempo transcurrido, el tiempo de planificación y otros metadatos de una consulta específica.

```
SELECT user_id,
       query_id,
       transaction_id,
       session_id,
       status,
       trim(database_name) AS database_name,
       start_time,
       end_time,
       result_cache_hit,
       elapsed_time,
       queue_time,
       execution_time,
       planning_time,
       trim(query_text) as query_text
FROM sys_query_history
WHERE query_id = 3093;
```

Resultados de ejemplo.

```
user_id | query_id | transaction_id | session_id |   status   | database_name |
   start_time         |          end_time          | result_cache_hit | elapsed_time |
 queue_time | execution_time | planning_time | query_text
--------+----------+----------------+------------+------------
+---------------+----------------------------+----------------------------
+------------------+-------------+-----------+----------------+--------------
+-------------------------------------
    106 |     3093 |          11759 | 1073750146 | success    | dev           |
 2023-03-16 16:53:17.840214 | 2023-03-16 16:53:18.106588 | f                |
 266374 |          0 |         105725 |        136589 | select count(*) from item;
```

La siguiente consulta muestra las 10 consultas SELECT más recientes.

```
SELECT query_id,
       transaction_id,
       session_id,
       start_time,
       elapsed_time,
       queue_time,
       execution_time,
       returned_rows,
       returned_bytes
FROM sys_query_history
WHERE query_type = 'SELECT'
ORDER BY start_time DESC limit 10;
```

Resultados de ejemplo.

```
 query_id | transaction_id | session_id |         start_time         | elapsed_time |
 queue_time | execution_time | returned_rows | returned_bytes
----------+----------------+------------+----------------------------+--------------
+------------+----------------+---------------+----------------
   526532 |          61093 | 1073840313 | 2022-02-09 04:43:24.149603 |       520571 |
        0 |         481293 |             1 |           3794
   526520 |          60850 | 1073840313 | 2022-02-09 04:38:27.24875  |       635957 |
        0 |         596601 |             1 |           3679
   526508 |          60803 | 1073840313 | 2022-02-09 04:37:51.118835 |       563882 |
        0 |         503135 |             5 |          17216
   526505 |          60763 | 1073840313 | 2022-02-09 04:36:48.636224 |       649337 |
        0 |         589823 |             1 |            652
```

```
   526478 |          60730 | 1073840313 | 2022-02-09 04:36:11.741471 |     14611321 |
        0 |       14544058 |             0 |              0
   526467 |          60636 | 1073840313 | 2022-02-09 04:34:11.91463  |     16711367 |
        0 |       16633767 |             1 |            575
   511617 |         617946 | 1074009948 | 2022-01-20 06:21:54.44481  |      9937090 |
        0 |        9899271 |           100 |          12500
   511603 |         617941 | 1074259415 | 2022-01-20 06:21:45.71744  |      8065081 |
        0 |        7582500 |           100 |           8889
   511595 |         617935 | 1074128320 | 2022-01-20 06:21:44.030876 |      1051270 |
        0 |        1014879 |             1 |             72
   511584 |         617931 | 1074030019 | 2022-01-20 06:21:42.764088 |       609033 |
        0 |         485887 |           100 |           8438
```

La siguiente consulta muestra el recuento diario de consultas select y el tiempo transcurrido promedio de consultas.

```
SELECT date_trunc('day',start_time) AS exec_day,
       status,
       COUNT(*) AS query_cnt,
       AVG(datediff (microsecond,start_time,end_time)) AS elapsed_avg
FROM sys_query_history
WHERE query_type = 'SELECT'
AND start_time >= '2022-01-14'
AND start_time <= '2022-01-18'
GROUP BY exec_day,
         status
ORDER BY exec_day,
         status;
```

Resultados de ejemplo.

```
      exec_day       | status  | query_cnt | elapsed_avg
---------------------+---------+-----------+------------
 2022-01-14 00:00:00 | success |      5253 |  56608048
 2022-01-15 00:00:00 | success |      7004 |  56995017
 2022-01-16 00:00:00 | success |      5253 |  57016363
 2022-01-17 00:00:00 | success |      5309 |  55236784
 2022-01-18 00:00:00 | success |      8092 |  54355124
```

La siguiente consulta muestra el rendimiento diario de tiempo transcurrido de consulta.

```
SELECT distinct date_trunc('day',start_time) AS exec_day,
```

```
       query_count.cnt AS query_count,
       Percentile_cont(0.5) within group(ORDER BY elapsed_time) OVER (PARTITION BY
 exec_day) AS P50_runtime,
       Percentile_cont(0.8) within group(ORDER BY elapsed_time) OVER (PARTITION BY
 exec_day) AS P80_runtime,
       Percentile_cont(0.9) within group(ORDER BY elapsed_time) OVER (PARTITION BY
 exec_day) AS P90_runtime,
       Percentile_cont(0.99) within group(ORDER BY elapsed_time) OVER (PARTITION BY
 exec_day) AS P99_runtime,
       Percentile_cont(1.0) within group(ORDER BY elapsed_time) OVER (PARTITION BY
 exec_day) AS max_runtime
FROM sys_query_history
LEFT JOIN (SELECT  date_trunc('day',start_time) AS day, count(*) cnt
           FROM sys_query_history
           WHERE query_type = 'SELECT'
           GROUP by 1) query_count
ON date_trunc('day',start_time) = query_count.day
WHERE query_type = 'SELECT'
ORDER BY exec_day;
```

Resultados de ejemplo.

```
      exec_day       | query_count | p50_runtime | p80_runtime | p90_runtime |
 p99_runtime  | max_runtime
--------------------+-------------+-------------+-------------+-------------
+-------------+-------------
 2022-01-14 00:00:00 |        5253 |  16816922.0 |  69525096.0 | 158524917.8 |
 486322477.52 | 1582078873.0
 2022-01-15 00:00:00 |        7004 |  15896130.5 |  71058707.0 | 164314568.9 |
 500331542.07 | 1696344792.0
 2022-01-16 00:00:00 |        5253 |  15750451.0 |  72037082.2 | 159513733.4 |
 480372059.24 | 1594793766.0
 2022-01-17 00:00:00 |        5309 |  15394513.0 |  68881393.2 | 160254700.0 |
 493372245.84 | 1521758640.0
 2022-01-18 00:00:00 |        8092 |  15575286.5 |  68485955.4 | 154559572.5 |
 463552685.39 | 1542783444.0
 2022-01-19 00:00:00 |        5860 |  16648747.0 |  72470482.6 | 166485138.2 |
 492038228.67 | 1693483241.0
 2022-01-20 00:00:00 |        1751 |  15422072.0 |  69686381.0 | 162315385.0 |
 497066615.00 | 1439319739.0
 2022-02-09 00:00:00 |          13 |   6382812.0 |  17616161.6 |  21197988.4 |
 23021343.84 |   23168439.0
```

La siguiente consulta muestra la distribución de tipos de consulta.

```
SELECT query_type,
       COUNT(*) AS query_count
FROM sys_query_history
GROUP BY query_type
ORDER BY query_count DESC;
```

Resultados de ejemplo.

```
 query_type | query_count
------------+-------------
 UTILITY    |      134486
 SELECT     |       38537
 DDL        |        4832
 OTHER      |         768
 LOAD       |         768
 CTAS       |         748
 COMMAND    |          92
```

En el siguiente ejemplo, se muestra la diferencia en los resultados del hash de consulta entre varias consultas. Observe las siguientes consultas:

```
CREATE TABLE test_table (col1 INT);

INSERT INTO test_table VALUES (1),(2);

SELECT * FROM test_table;

SELECT * FROM test_table;

SELECT col1 FROM test_table;

SELECT * FROM test_table WHERE col1=1;

SELECT * FROM test_table WHERE col1=2;

SELECT query_id, TRIM(user_query_hash) AS user_query_hash, TRIM(generic_query_hash)
 AS generic_query_hash, TRIM(query_text) AS text FROM sys_query_history ORDER BY
 start_time
DESC LIMIT 10;
```

A continuación, se muestra un ejemplo de salida:

```
query_id | user_query_hash | generic_query_hash | text
---------+-----------------+--------------------+----------
24723049 | oPuFtjEPLTs=    | oPuFtjEPLTs=       | select query_id,
 trim(user_query_hash) as user_query_hash, trim(generic_query_hash) as
 generic_query_hash, query_hash_version, trim(query_text) as text from
 sys_query_history order by start_time\r\ndesc limit 20
24723045 | Gw2Kwdd8m2I=    | IwfRu8/XAKI=       | select * from test_table where col1=2
 limit 100
24723041 | LNw2vx0GDXo=    | IwfRu8/XAKI=       | select * from test_table where col1=1
 limit 100
24723036 | H+qep/c82Y8=    | H+qep/c82Y8=       | select col1 from test_table limit 100
24723033 | H+qep/c82Y8=    | H+qep/c82Y8=       | select * from test_table limit 100
24723029 | H+qep/c82Y8=    | H+qep/c82Y8=       | select * from test_table limit 100
24723023 | 50sirx9E1hU=    | uO36Z1a/QYs=       | insert into test_table values (1),(2)
24723021 | YSVnlivZHeo=    | YSVnlivZHeo=       | create table test_table (col1 int)
```

`SELECT * FROM test_table;` y `SELECT col1 FROM test_table;` tienen el mismo valor de user_query_hash, ya que test_table solo tiene una columna. `SELECT * FROM test_table WHERE col1=1;` y `SELECT * FROM test_table WHERE col1=2;` tienen valores de user_query_hash diferentes, pero valores de generic_query_hash idénticos, ya que las dos consultas son idénticas fuera de los literales de consulta 1 y 2.

## SYS_QUERY_TEXT

Utilice SYS_QUERY_TEXT para ver el texto de todas las consultas. Cada fila representa el texto de las consultas de hasta 4000 caracteres que comienzan por el número de secuencia 0. Cuando la instrucción de consulta contiene más de 4000 caracteres, se registran filas adicionales para la instrucción al incrementar el número de secuencia de cada fila. Esta vista registra todo el texto de las consultas de los usuarios como las consultas de DDL, de utilidades, de Amazon Redshift y únicas del nodo principal.

SYS_QUERY_TEXT es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | El identificador del usuario que envió la consulta. |
| query_id | bigint | Identificador de la consulta. |
| transaction_id | bigint | El identificador de la transacción asociado a la instrucción. |
| session_id | entero | El identificador del proceso de la sesión que ejecuta la consulta. |
| start_time | marca de tiempo | La hora cuando empieza la consulta. |
| sequence | entero | Cuando una única instrucción tiene más de 4000 caracteres, se registran filas adicionales para la instrucción. Secuencia 0 es la primera fila, 1 es la segunda fila y, así, sucesivamente. |
| texto | carácter (4000) | El texto de la consulta SQL que está en incrementos de 4000 caracteres. Este campo puede contener caracteres especiales, como una barra oblicua inversa (\) y una nueva línea (\n). |

## Consultas de ejemplo

La siguiente consulta devuelve consultas en ejecución y en cola.

```
SELECT user_id,
 query_id,
 transaction_id,
 session_id, start_time,
 sequence, trim(text) as text from sys_query_text
 ORDER BY sequence;
```

Resultados de ejemplo.

```
 user_id | query_id | transaction_id | session_id |        start_time          |
 sequence |                                                        text
--------+----------+----------------+------------+----------------------------
+----------
+-----------------------------------------------------------------------------------------------------------
   100  |     4    |       1396     | 1073750220 | 2023-04-28 16:44:55.887184 |
   0    | SELECT trim(text) as text, sequence FROM sys_query_text WHERE query_id =
 pg_last_query_id() AND user_id > 1 AND start
_time > '2023-04-28 16:44:55.922705+00:00'::timestamp order by sequence;
```

La siguiente consulta devuelve los permisos concedidos o revocados a los grupos de la base de datos.

```
SELECT
    SPLIT_PART(text, ' ', 1) as grantrevoke,
    SPLIT_PART((SUBSTRING(text, STRPOS(UPPER(text), 'GROUP'))), ' ', 2) as group,
    SPLIT_PART((SUBSTRING(text, STRPOS(UPPER(text), ' '))), 'ON', 1) as type,
    SPLIT_PART((SUBSTRING(text, STRPOS(UPPER(text), 'ON'))), ' ', 2) || ' ' ||
 SPLIT_PART((SUBSTRING(text, STRPOS(UPPER(text), 'ON'))), ' ', 3) as entity
FROM SYS_QUERY_TEXT
WHERE (text LIKE 'GRANT%' OR text LIKE 'REVOKE%') AND text LIKE '%GROUP%';

+-------------+----------+--------+----------+
| grantrevoke |  group   |  type  |  entity  |
+-------------+----------+--------+----------+
| GRANT       | bi_group | SELECT | TABLE t1 |
| GRANT       | bi_group | SELECT | TABLE t1 |
| GRANT       | bi_group | SELECT | TABLE t1 |
| GRANT       | bi_group | USAGE  | TABLE t1 |
| GRANT       | bi_group | SELECT | TABLE t1 |
| GRANT       | bi_group | SELECT | TABLE t1 |
+-------------+----------+--------+----------+
```

# SYS_RESTORE_LOG

Utilice SYS_RESTORE_LOG para monitorear la evolución de la migración de cada tabla del clúster durante un cambio de tamaño clásico hacia los nodos RA3. Captura el rendimiento histórico de la migración de datos durante la operación de cambio de tamaño. Para obtener más información sobre el cambio de tamaño clásico hacia los nodos RA3, consulte Cambio de tamaño clásico.

SYS_RESTORE_STATE solo es visible para los superusuarios.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| event_time | marca de tiempo | Es una marca de tiempo que indica cuándo se registra la entrada del registro. |
| database_name | char(128) | El nombre de la base de datos. |
| schema_name | char(128) | El nombre del esquema. |
| table_name | char(128) | El nombre de la tabla. |
| table_id | entero | ID de la tabla. |
| acción | char(128) | Acción realizada en el momento de la entrada. Los valores pueden ser: migración iniciada, punto de comprobación, reanudada, completada, cancelada o restablecida. |
| table_size | long | Es el tamaño de la tabla. |
| total_data_processed | long | Es el tamaño de los datos en MB procesados hasta el momento para la tabla. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| delta_data_processed | long | Es el tamaño de los datos procesados desde la última actualización de event_time en MB. Esto ayuda a determina r qué cantidad de datos se ha procesado desde el intervalo de tiempo registrado anteriorm ente. Se puede comparar con el valor de table_size para hacerse una idea de la rapidez con la que se procesan los datos. |
| message | char(512) | Explicación detallada del valor en la columna de acción. |
| redistribution_type | char(32) | Es el tipo de redistribución de la tabla. Puede ser una conversión KEY o una tarea de reequilibrio EVEN. Para obtener más información sobre los estilos de distribuc ión, consulte Estilos de distribución. |

## Consultas de ejemplo

La siguiente consulta calcula el rendimiento del procesamiento de datos mediante SYS_RESTORE_LOG.

```
SELECT
    ROUND(sum(delta_data_processed) / 1024.0, 2) as data_processed_gb,
    ROUND(datediff(sec, min(event_time), max(event_time)) / 3600.0, 2) as duration_hr,
    ROUND(data_processed_gb/duration_hr, 2) as throughput_gb_per_hr
from sys_restore_log;
```

Resultados de ejemplo.

```
 data_processed_gb | duration_hr | throughput_gb_per_hr
-------------------+-------------+----------------------
              0.91 |        8.37 |                 0.11
(1 row)
```

La siguiente consulta muestra todos los tipos de redistribución.

```
SELECT * from sys_restore_log ORDER BY event_time;
```

```
 database_name |     schema_name      |      table_name      | table_id |
 action             | total_data_processed | delta_data_processed |         event_time
        | table_size | message |   redistribution_type
---------------+----------------------+----------------------+----------
+----------------------------+----------------------+----------------------
+----------------------------+------------+---------+--------------------------
 dev           | schemaaaa877096d844d | customer_key         |   106424 |
 Redistribution started      |                    0 |                      | 2024-01-05
 02:18:00.744977 |        325 |         | Restore Distkey Table
 dev           | schemaaaa877096d844d | dp30907_t2_autokey   |   106430 |
 Redistribution started      |                    0 |                      | 2024-01-05
 02:18:02.756675 |         90 |         | Restore Distkey Table
 dev           | schemaaaa877096d844d | dp30907_t2_autokey   |   106430 |
 Redistribution completed    |                   90 |                   90 | 2024-01-05
 02:23:30.643718 |         90 |         | Restore Distkey Table
 dev           | schemaaaa877096d844d | customer_key         |   106424 |
 Redistribution completed    |                  325 |                  325 | 2024-01-05
 02:23:45.998249 |        325 |         | Restore Distkey Table
 dev           | schemaaaa877096d844d | dp30907_t1_even      |   106428 |
 Redistribution started      |                    0 |                      | 2024-01-05
 02:23:46.083849 |         30 |         | Rebalance Disteven Table
 dev           | schemaaaa877096d844d | dp30907_t5_auto_even |   106436 |
 Redistribution started      |                    0 |                      | 2024-01-05
 02:23:46.855728 |         45 |         | Rebalance Disteven Table
 dev           | schemaaaa877096d844d | dp30907_t5_auto_even |   106436 |
 Redistribution completed    |                   45 |                   45 | 2024-01-05
 02:24:16.343029 |         45 |         | Rebalance Disteven Table
 dev           | schemaaaa877096d844d | dp30907_t1_even      |   106428 |
 Redistribution completed    |                   30 |                   30 | 2024-01-05
 02:24:20.584703 |         30 |         | Rebalance Disteven Table
```

```
 dev           | schemaefd028a2a48a4c | customer_even        |   130512 |
 Redistribution started      |                    0 |                      | 2024-01-05
 04:54:55.641741 |        190 |         | Restore Disteven Table
 dev           | schemaefd028a2a48a4c | customer_even        |   130512 |
 Redistribution checkpointed |     29.4342113157737 |     29.4342113157737 | 2024-01-05
 04:55:04.770696 |        190 |         | Restore Disteven Table
(8 rows)
```

# SYS_RESTORE_STATE

Utilice SYS_RESTORE_STATE para monitorear la evolución de la migración de cada tabla durante un cambio de tamaño clásico. Esto es aplicable específicamente cuando el tipo de nodo de destino es RA3. Para obtener más información sobre el cambio de tamaño clásico hacia los nodos RA3, consulte Cambio de tamaño clásico.

SYS_RESTORE_STATE solo es visible para los superusuarios. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | El identificador del usuario que envió la consulta. |
| database_name | char(64) | Nombre de la base de datos de la tabla. |
| schema_id | entero | ID del esquema de la tabla. |
| table_id | entero | ID de la tabla. |
| table_name | char(128) | El nombre de la tabla. |
| redistribution_status | char(128) | Estado del progreso de redistribución de la tabla. Los posibles valores son `Completed` , `In progress` y `Pending`. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| percentage_redistributed | float | Porcentaje del progreso de la redistribución de la tabla. Los valores posibles oscilan entre 0 y 100 %. Por ejemplo, un valor de 25 indica que se ha redistribuido el 25 % de los datos. |
| redistribution_type | char(32) | Es el tipo de redistribución de la tabla. Puede ser una conversión KEY o una tarea de reequilibrio EVEN. Para obtener más información sobre los estilos de distribuc ión, consulte Estilos de distribución. |

## Consultas de ejemplo

La siguiente consulta devuelve los registros para las consultas en ejecución y en cola.

```
SELECT * FROM sys_restore_state;
```

Resultados de ejemplo.

```
 userid | database_name | schema_id | table_id |   table_name   | redistribution_status
 | precentage_redistributed |   redistribution_type
--------+---------------+-----------+----------+----------------
+----------------------+--------------------------+------------------------
    1   |     test1     |   124865  |  124878  | customer_key_4 |         Pending
 |      0                 |  Rebalance Disteven Table
    1   |      dev      |   124865  |  124874  | customer_key_3 |         Pending
 |      0                 |  Rebalance Disteven Table
    1   |      dev      |   124865  |  124870  | customer_key_2 |        Completed
 |     100                |  Rebalance Disteven Table
```

```
    1   |      dev      |   124865  |  124866  | customer_key_1 |       In progress
 |     13.52               |  Restore Distkey Table
```

A continuación se muestra el estado del procesamiento de datos.

```
SELECT
    redistribution_status, ROUND(SUM(block_count) / 1024.0, 2) AS total_size_gb
FROM sys_restore_state sys inner join stv_tbl_perm stv
    on sys.table_id = stv.id
GROUP BY sys.redistribution_status;
```

Resultados de ejemplo.

```
 redistribution_status | total_size_gb
-----------------------+---------------
 Completed             |          0.07
 Pending               |          0.71
 In progress           |          0.20
(3 rows)
```

# SYS_SCHEMA_QUOTA_VIOLATIONS

Registra la incidencia, marca temporal, ID y otros datos útiles cuando se excede una cuota de esquema. Esta tabla del sistema es una traducción de STL_SCHEMA_QUOTA_VIOLATIONS.

r_SYS_SCHEMA_QUOTA_VIOLATIONS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| owner_id | entero | El ID del propietario del esquema. |
| user_id | entero | ID del usuario que generó la entrada. |
| transacti on_id | bigint | El ID de la transacción asociado a la instrucción. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| session_id | entero | El ID del proceso asociado a la instrucción. |
| schema_id | entero | El espacio de nombres o el ID de esquema. |
| schema_name | carácter (128) | El espacio de nombres o el nombre del esquema. |
| cuota | entero | La cantidad de espacio en disco (en MB) que puede utilizar el esquema. |
| disk_usage | entero | El espacio en disco (en MB) que utiliza actualmente el esquema. |
| record_time | TIMESTAMP sin zona horaria | La hora en que se produjo la infracción. |

### Consultas de ejemplo

En la siguiente consulta se muestra el resultado de la infracción de cuota:

```
SELECT user_id, TRIM(schema_name) "schema_name", quota, disk_usage, record_time FROM
sys_schema_quota_violations WHERE SCHEMA_NAME = 'sales_schema' ORDER BY timestamp DESC;
```

Esta consulta devuelve el siguiente resultado de ejemplo para el esquema especificado:

```
user_id| schema_name  | quota | disk_usage | record_time
-------+--------------+-------+------------+----------------------------
104    | sales_schema | 2048  | 2798       | 2020-04-20 20:09:25.494723
(1 row)
```

## SYS_SERVERLESS_USAGE

Utilice SYS_SERVERLESS_USAGE para ver los detalles del uso de los recursos de Amazon Redshift sin servidor. Esta vista del sistema no se aplica a los clústeres aprovisionados de Amazon Redshift.

Esta vista contiene el resumen de uso sin servidor, incluidas la capacidad de computación utilizada para procesar consultas y la cantidad de almacenamiento administrado de Amazon Redshift que se utiliza con una precisión de 1 minuto. La capacidad de computación se mide en unidades de procesamiento Redshift (RPU) y se dosifica para las cargas de trabajo que se ejecutan en RPU por segundo. Las RPU se utilizan para procesar consultas sobre los datos cargados en el almacenamiento de datos, que se consultan desde un lago de datos de Amazon S3 o a los que se obtiene acceso desde las bases de datos operativas mediante una consulta federada. Amazon Redshift sin servidor conserva la información en SYS_SERVERLESS_USAGE durante 7 días.

Para ver ejemplos de facturación de costos de computación, consulte Facturación de Amazon Redshift sin servidor.

SYS_SERVERLESS_USAGE es visible para todos los superusuarios. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| start_time | marca de tiempo | La hora de inicio del intervalo. |
| end_time | marca de tiempo | La hora en que se completó el intervalo. |
| compute_seconds | double precision | La cantidad de segundos de unidad de computación acumulada (RPU) que se ha consumido durante este intervalo de tiempo. Este valor representa la capacidad de RPU base asignada a la cuenta. |
| compute_capacity | double precision | El número medio de unidades de computación (unidades de procesamiento de Redshift o RPU) asignadas durante este intervalo de tiempo. |

| Nombre de la columna | Tipo de datos: | Descripción |
| --- | --- | --- |
| | | El valor de compute_capacity puede modificarse de forma dinámica. |
| data_storage | entero | El espacio medio de almacenamiento de datos en MB utilizado durante este intervalo de tiempo.<br><br>El almacenamiento de datos utilizados puede modificarse de forma dinámica a medida que se cargan o eliminan datos de la base de datos. |
| cross_region_transferred_data | entero | Los datos acumulados transferidos para recursos compartidos de datos entre regiones en bytes durante este intervalo de tiempo. |

| Nombre de la columna | Tipo de datos: | Descripción |
| --- | --- | --- |
| charged_seconds | entero | La cantidad de segundos de unidad de computación acumulada (RPU) que se ha cargado durante este intervalo de tiempo. Se computa una vez finalizadas las transacciones, por lo que puede ser 0 mientras se ejecuta una transacción. Utilice charged_seconds para calcular el costo de un grupo de trabajo de Amazon Redshift sin servidor. Este valor representa la capacidad de RPU asignada para el grupo de trabajo de Amazon Redshift sin servidor. |

## Notas de uso

- Hay situaciones en las que compute_seconds es 0 pero charged_seconds es mayor que 0 o viceversa. Este es un comportamiento normal debido a la forma en que se graban los datos en la vista del sistema. Para obtener una representación más precisa de los detalles del uso sin servidor, recomendamos agregar los datos.

## Ejemplo

Para obtener el total de cargos por horas de RPU utilizadas durante un intervalo de tiempo mediante la consulta charged_seconds, ejecute la siguiente consulta:

```
select trunc(start_time) "Day",
(sum(charged_seconds)/3600::double precision) * <Price for 1 RPU> as cost_incurred
from sys_serverless_usage
group by 1
order by 1
```

Tenga en cuenta que puede haber tiempo de inactividad durante el intervalo. El tiempo inactivo no se suma a las RPU consumidas.

# SYS_SESSION_HISTORY

Utilice SYS_SESSION_HISTORY para ver información sobre las sesiones activas actuales y el historial de sesiones.

SYS_SESSION_HISTORY es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | char(50) | Identificador del usuario que generó la entrada. |
| session_id | entero | El ID de la sesión asociada a la instrucción. |
| database_name | char(50) | El nombre de la base de datos. |
| status | char | El estado de la sesión. Los posibles valores son `active`, `timed out` y `closed`. |
| session_timeout | entero | El tiempo máximo en segundos durante el cual una sesión permanece inactiva o en reposo antes de que se cierre. El 0 indica que no se establece tiempo de espera. |
| start_time | marca de tiempo | La marca temporal en la que se estableció la conexión. |
| end_time | marca de tiempo | La marca temporal en la que se detuvo la conexión. |

### Ejemplo

A continuación se muestra un ejemplo de salida de SYS_SESSION_HISTORY.

```
select * from sys_session_history;
```

```
 user_id | session_id |   database_name  | status | session_timeout |
 start_time         |          end_time
---------+------------+------------------+--------+-----------------
+---------------------------+---------------------------
       1 | 1073971370 | dev              | closed |        0        | 2023-07-17
 15:50:12.030104 | 2023-07-17 15:50:12.123218
       1 | 1073979694 | dev              | closed |        0        | 2023-07-17
 15:50:24.117947 | 2023-07-17 15:50:24.131859
       1 | 1073873049 | dev              | closed |        0        | 2023-07-17
 15:49:29.067398 | 2023-07-17 15:49:29.070294
       1 | 1073873086 | database18127a4a | closed |        0        | 2023-07-17
 15:49:29.119018 | 2023-07-17 15:49:29.125925
       1 | 1073832112 | dev              | closed |        0        | 2023-07-17
 15:49:29.164688 | 2023-07-17 15:49:29.179631
       1 | 1073987697 | dev              | closed |        0        | 2023-07-17
 15:49:29.26749  | 2023-07-17 15:49:29.273034
       1 | 1073922429 | dev              | closed |        0        | 2023-07-17
 15:49:33.35315  | 2023-07-17 15:49:33.367499
       1 | 1073766783 | dev              | closed |        0        | 2023-07-17
 15:49:45.38237  | 2023-07-17 15:49:45.396902
       1 | 1073807506 | dev              | active |        0        | 2023-07-17
 15:51:48        |
```

## SYS_SPATIAL_SIMPLIFY

Puede consultar la vista del sistema SYS_SPATIAL_SIMPLIFY para obtener información acerca de los objetos de geometría espacial simplificada mediante el comando COPY. Cuando utiliza COPY en un shapefile, puede especificar las opciones de ingesta SIMPLIFY `tolerance`, SIMPLIFY AUTO y SIMPLIFY AUTO `max_tolerance`. El resultado de la simplificación se resume en la vista de sistema SYS_SPATIAL_SIMPLIFY.

Cuando se configura SIMPLIFY AUTO `max_tolerance`, esta vista contiene una fila para cada geometría que excedió el tamaño máximo. Cuando se configura SIMPLIFY `tolerance`, se almacena una fila para toda la operación COPY. Esta fila referencia el ID de consulta COPY y la tolerancia de simplificación especificada.

Para obtener más información sobre cómo cargar un shapefile, consulte Carga de un shapefile en Amazon Redshift.

SYS_SPATIAL_SIMPLIFY es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| query_id | bigint | El ID de la consulta (comando COPY) que generó esta fila. |
| line_number | bigint | Cuando se especifica la opción COPY `SIMPLIFY AUTO`, este valor es el número de registro del registro simplificado en el shapefile. |
| maximum_tolerance | double precision | El valor de tolerancia de distancia especificado en el comando COPY. Este es el valor de tolerancia máxima mediante la opción `SIMPLIFY AUTO` o el valor de tolerancia fijo mediante la opción `SIMPLIFY`. |
| inicial_size | bigint | El tamaño en bytes del valor de datos `GEOMETRY` antes de la simplificación. |
| simplified | char(1) | Cuando se especifica la opción COPY `SIMPLIFY AUTO`, `t` si la geometría se simplificó correctamente, o, de lo contrario, `f`. Es posible que la geometría no se simplifique correctamente si después de la simplificación con la tolerancia máxima dada su tamaño sigue siendo mayor que el tamaño máximo de la geometría. |
| final_size | bigint | Cuando se especifica la opción COPY `SIMPLIFY AUTO`, este es el tamaño en bytes de la geometría después de la simplificación. |
| final_tolerance | double precision | Se eligió la tolerancia final para la simplificación. |

## Consulta de ejemplo

La siguiente consulta devuelve la lista de registros que COPY simplificó.

```
SELECT * FROM sys_spatial_simplify;
```

```
 query_id | line_number | maximum_tolerance | initial_size | simplified | final_size |
  final_tolerance
----------+-------------+-------------------+--------------+------------+------------
+----------------------
    20    |     1184704 |                -1 |      1513736 | t          |    1008808 |
  1.276386653895e-05
    20    |     1664115 |                -1 |      1233456 | t          |    1023584 |
 6.11707814796635e-06
```

# SYS_STREAM_SCAN_ERRORS

Registra los errores de los registros cargados mediante la ingesta de streaming.

SYS_STREAM_SCAN_ERRORS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| external_schema_name | character (128) | Nombre del esquema de Kinesis o del tema de Amazon MSK. Distingue entre mayúsculas y minúsculas. |
| stream_name | character (255) | Nombre del flujo o tema. Distingue entre mayúsculas y minúsculas. |
| mv_name | character (128) | Nombre de la vista materializada asociada. Está vacío si no hay ninguna. Distingue entre mayúsculas y minúsculas. |
| transaction_id | bigint | El ID de la transacción. |
| query_id | bigint | El ID de la consulta. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| stream_timestamp_type | character (1) | El tipo de marca temporal de flujo. Distingue entre mayúsculas y minúsculas. |
| stream_timestamp | TIMESTAM sin zona horaria | Hora en que llegó el registro. |
| record_time | TIMESTAM sin zona horaria | Hora en que se registró el mensaje de error. |
| partition_id | character (128) | Identificador de la partición. Distingue entre mayúsculas y minúsculas. |
| position | character (128) | Posición del registro. Se corresponde con el número de secuencia en Kinesis o el desplazamiento en Amazon MSK. Distingue entre mayúsculas y minúsculas. |
| error_code | entero | Código de error. |
| error_reason | character (128) | Motivo del error. Distingue entre mayúsculas y minúsculas. |

## SYS_STREAM_SCAN_STATES

Registra los estados de exploración de los registros cargados mediante la ingesta de streaming.

SYS_STREAM_SCAN_STATES es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| external_schema_name | character (128) | El nombre del esquema externo. Distingue entre mayúsculas y minúsculas. |
| stream_name | character (255) | El nombre de la transmisión. Distingue entre mayúsculas y minúsculas. |
| mv_name | character (128) | Nombre de la vista materializada asociada. Está vacío si no hay ninguna. Distingue entre mayúsculas y minúsculas. |
| transaction_id | bigint | El ID de la transacción. |
| query_id | bigint | El ID de la consulta. |
| record_time | TIMESTAM sin zona horaria | Hora en que se registraron los datos. |
| partition_id | character (128) | Identificador de la partición. Distingue entre mayúsculas y minúsculas. |
| latest_position | character (128) | Posición del último registro leído en el lote. Se corresponde con el número de secuencia en Kinesis o el desplazamiento en Amazon MSK. Distingue entre mayúsculas y minúsculas. |
| scanned_rows | bigint | Número de registros que se han explorado en el lote. |
| skipped_rows | bigint | Número de registros que se han omitido en el lote. |
| scanned_bytes | bigint | Número de bytes que se han explorado en el lote. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| stream_record_time_min | TIMESTAM sin zona horaria | Hora de llegada de Kinesis o Amazon MSK para el registro más antiguo del lote. |
| stream_record_time_max | TIMESTAM sin zona horaria | Hora de llegada de Kinesis o Amazon MSK para el registro más reciente del lote. |

En la siguiente consulta se muestran los datos de flujo y de tema para consultas específicas.

```
select
 query_id,mv_name::varchar,external_schema_name::varchar,stream_name::varchar,sum(scanned_rows)
 total_records,
sum(scanned_bytes) total_bytes from sys_stream_scan_states where query in
 (5401180,8601939) group by 1,2,3,4;

  query_id  |    mv_name     | external_schema_name |   stream_name   | total_records |
  total_bytes
------------+----------------+----------------------+-----------------+---------------
+----------------
 5401180    | kinesistest    | kinesis              | kinesisstream   |    1493255696 |
 3209006490704
 8601939    | msktest        | msk                  | mskstream       |      14677023 |
   31056580668
(2 rows)
```

## SYS_TRANSACTION_HISTORY

Utilice SYS_TRANSACTION_HISTORY para ver los detalles de una transacción al realizar el seguimiento de una consulta.

SYS_TRANSACTION_HISTORY solo es visible para los superusuarios. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | ID del usuario que generó la entrada. |
| transaction_id | bigint | El ID de la transacción. |
| isolation_level | texto | El nivel de aislamiento de la transacción. Los posibles valores son `Serializable` y `Snapshot Isolation` . |
| status | texto | El estado de la transacción. Los estados posibles son `committed` y `rolledback` . |
| transaction_start_time | marca de tiempo | La hora de inicio de la transacción. |
| commit_start_time | marca de tiempo | La hora de inicio de la confirmación. |
| commit_end_time | marca de tiempo | La hora de finalización de la confirmación. |
| blocks_committed | bigint | El número de bloques que tuvieron que escribirse como parte de esta confirmación. |
| undo_transaction_id | bigint | El ID de la transacción deshecha si se ha deshecho alguna transacción. De lo contrario, este valor es -1. |

## Consultas de ejemplo

```
select * from sys_transaction_history order by transaction_start_time desc;

 user_id | transaction_id | isolation_level |   status   |   transaction_start_time
   |     commit_start_time      |      commit_end_time       | blocks_committed |
 undo_transaction_id
---------+----------------+-----------------+------------+--------------------------
+---------------------------+---------------------------+------------------
+---------------------
     100 |           1310 | Serializable    | committed  | 2023-08-27 21:03:11.822205 |
 2023-08-28 21:03:11.825287 | 2023-08-28 21:03:11.854883 |               17 |
       -1
     101 |           1345 | Serializable    | committed  | 2023-08-27 21:03:12.000278 |
 2023-08-28 21:03:12.003736 | 2023-08-28 21:03:12.030061 |               17 |
       -1
     102 |           1367 | Serializable    | committed  | 2023-08-27 21:03:12.1532   |
 2023-08-28 21:03:12.156124 | 2023-08-28 21:03:12.186468 |               17 |
       -1
     100 |           1370 | Serializable    | committed  | 2023-08-27 21:03:12.199316 |
 2023-08-28 21:03:12.204854 | 2023-08-28 21:03:12.238186 |               24 |
       -1
     100 |           1408 | Serializable    | committed  | 2023-08-27 21:03:53.891107 |
 2023-08-28 21:03:53.894825 | 2023-08-28 21:03:53.927465 |               17 |
       -1
     100 |           1409 | Serializable    | rolledback | 2023-08-27 21:03:53.936431 |
 2000-01-01 00:00:00        | 2023-08-28 21:04:08.712532 |                0 |
      1409
     101 |           1415 | Serializable    | committed  | 2023-08-27 21:04:24.283188 |
 2023-08-28 21:04:24.289196 | 2023-08-28 21:04:24.374318 |               25 |
       -1
     101 |           1416 | Serializable    | committed  | 2023-08-27 21:04:24.38818  |
 2023-08-28 21:04:24.391688 | 2023-08-28 21:04:24.415135 |               17 |
       -1
     100 |           1417 | Serializable    | rolledback | 2023-08-27 21:04:24.424252 |
 2000-01-01 00:00:00        | 2023-08-28 21:04:28.354826 |                0 |
      1417
     101 |           1418 | Serializable    | rolledback | 2023-08-27 21:04:24.425195 |
 2000-01-01 00:00:00        | 2023-08-28 21:04:28.680355 |                0 |
      1418
     100 |           1420 | Serializable    | committed  | 2023-08-27 21:04:28.697607 |
 2023-08-28 21:04:28.702374 | 2023-08-28 21:04:28.735541 |               23 |
       -1
```

```
    101 |           1421 | Serializable    | committed  | 2023-08-27 21:04:28.744854 |
2023-08-28 21:04:28.749344 | 2023-08-28 21:04:28.779029 |               23 |
       -1
    101 |           1423 | Serializable    | committed  | 2023-08-27 21:04:28.78942  |
2023-08-28 21:04:28.791556 | 2023-08-28 21:04:28.817485 |               16 |
       -1
    100 |           1430 | Serializable    | committed  | 2023-08-27 21:04:28.917788 |
2023-08-28 21:04:28.919993 | 2023-08-28 21:04:28.944812 |               16 |
       -1
    102 |           1494 | Serializable    | committed  | 2023-08-27 21:04:37.029058 |
2023-08-28 21:04:37.033137 | 2023-08-28 21:04:37.062001 |               16 |
       -1
```

# SYS_UDF_LOG

Registra los mensajes de error y advertencia definidos por el sistema generados durante la ejecución de la función definida por el usuario (UDF).

SYS_UDF_LOG solo está visible para los superusuarios. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| query_id | bigint | Identificador de la consulta. |
| function_name | texto | El nombre de la función definida por el usuario. |
| record_time | marca de tiempo | La hora en la que se creó el registro. |
| sequence | entero | La secuencia de un único mensaje de registro. |
| message | texto | El texto del mensaje de registro. |

## Consultas de ejemplo

En el siguiente ejemplo, se muestra cómo las UDF administran los errores definidos por el sistema. El primer bloque muestra la definición para una función UDF que devuelve la inversa de un argumento. Si ejecuta la función y proporciona un 0 como argumento, la función devuelve un error. La última instrucción devuelve el mensaje de error registrado en SYS_UDF_LOG.

```
-- Create a function to find the inverse of a number.
CREATE OR REPLACE FUNCTION f_udf_inv(a int)

RETURNS float

IMMUTABLE AS $$return 1/a

$$ LANGUAGE plpythonu;

-- Run the function with 0 to create an error.
Select f_udf_inv(0);

-- Query SYS_UDF_LOG to view the message.
Select query_id, record_time, message::varchar from sys_udf_log;


query_id    |    record_time              |                message
----------+----------------------------
+-------------------------------------------------------
2211        | 2023-08-23 15:53:11.360538 |  ZeroDivisionError: integer division or
 modulo by zero line 2, in f_udf_inv\n return 1/a\n
```

En el siguiente ejemplo, se agrega el registro y un mensaje de advertencia a la UDF para que una operación de división en cero dé como resultado un mensaje de advertencia, en lugar de parar con un mensaje de error.

```
-- Create a function to find the inverse of a number and log a warning if you input 0.
CREATE OR REPLACE FUNCTION f_udf_inv_log(a int)
  RETURNS float IMMUTABLE
 AS $$
  import logging
  logger = logging.getLogger() #get root logger
  if a==0:
    logger.warning('You attempted to divide by zero.\nReturning zero instead of error.
\n')
```

```
    return 0
  else:
     return 1/a
$$ LANGUAGE plpythonu;

-- Run the function with 0 to trigger the warning.
Select f_udf_inv_log(0);

-- Query SYS_UDF_LOG to view the message.
Select query_id, record_time, message::varchar from sys_udf_log;

 query_id |        record_time         |                                        message
----------+----------------------------
+-----------------------------------------------------------------------------------
     0   | 2023-08-23 16:10:48.833503 | WARNING: You attempted to divide by zero.
\nReturning zero instead of error.\n
```

# SYS_UNLOAD_DETAIL

Utilice SYS_UNLOAD_TAIL para ver los detalles de la operación UNLOAD. Registra una fila por cada archivo creado por una instrucción UNLOAD. Por ejemplo, si una instrucción UNLOAD crea 12 archivos, SYS_UNLOAD_DETAIL contendrá 12 filas correspondientes.

SYS_UNLOAD_DETAIL es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | Identificador del usuario que generó la entrada. |
| query_id | entero | El identificador de consulta del comando UNLOAD. |
| session_id | entero | ID del proceso asociado a la instrucción de la consulta. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| transaction_id | bigint | ID de la transacción asociado a la instrucción de la consulta. |
| file_name | character (1280) | La ruta competa del objeto de Amazon S3 para el archivo. |
| start_time | marca de tiempo | La hora a la que comenzó la transacción. |
| end_time | marca de tiempo | La hora a la que se completó la transacción. |
| line_count | bigint | Cantidad de líneas (filas) descargadas en el archivo. |
| transfer_size | bigint | Número de bytes transferidos. |
| file_format | character (10) | Formato de los archivos descargados. |

## Consultas de ejemplo

La siguiente consulta muestra los detalles de la consulta descargada, incluidos el formato, las filas y el recuento de archivos del comando unload.

```
select query_id, substring(file_name, 0, 50), transfer_size, file_format from
 sys_unload_detail;
```

Resultados de ejemplo.

```
 query_id |                    substring                                | transfer_size
 | file_format
----------+-----------------------------------------------------------
+---------------+-------------
     9272 | s3://amzn-s3-demo-bucket/my_unload_doc_venue0000_part_00.gz  |
 395886 | Text
```

```
    9272 | s3://amzn-s3-demo-bucket/my_unload_doc_venue0001_part_00.gz  |
 406444 | Text
    9272 | s3://amzn-s3-demo-bucket/my_unload_doc_venue0002_part_00.gz  |
 409431 | Text
    9272 | s3://amzn-s3-demo-bucket/my_unload_doc_venue0003_part_00.gz  |
 403051 | Text
    9272 | s3://amzn-s3-demo-bucket/my_unload_doc_venue0004_part_00.gz  |
 413592 | Text
    9272 | s3://amzn-s3-demo-bucket/my_unload_doc_venue0005_part_00.gz  |
 395689 | Text
(6 rows)
```

# SYS_UNLOAD_HISTORY

Utilice SYS_UNLOAD_HISTORY para visualizar los detalles de los comandos UNLOAD. Cada una de las filas representa un comando UNLOAD con estadísticas acumuladas para algunos de los campos. Contiene tanto los comandos UNLOAD en ejecución como los que han finalizado.

SYS_UNLOAD_HISTORY es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | El identificador del usuario que envió la descarga. |
| query_id | bigint | El identificador de consulta del comando UNLOAD. |
| transaction_id | bigint | Identificador de la transacción. |
| session_id | entero | El identificador de proceso del proceso que ejecuta la descarga. |
| database_name | texto | El nombre de la base de datos a la que estaba conectado el |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | usuario cuando se emitió la operación. |
| status | texto | El estado del comando UNLOAD. Los valores válidos son `running`, `completed` , `aborted` y `unknown`. |
| start_time | marca de tiempo | La hora de inicio de la descarga. |
| end_time | marca de tiempo | La hora en que se completó la descarga. |
| duración | bigint | La cantidad de tiempo (microsegundos) empleados en el comando UNLOAD. |
| file_format | texto | El formato de los archivos de salida. |
| compression_type | texto | El tipo de compresión. |
| unloaded_location | texto | La ubicación de Amazon S3 de los archivos descargados. |
| unloaded_rows | bigint | La cantidad de filas. |
| unloaded_files_count | bigint | La cantidad de archivos del archivo de salida. |
| unloaded_files_size | bigint | El tamaño de archivos del archivo de salida. |
| error_message | texto | El mensaje de error del comando UNLOAD. |

### Consultas de ejemplo

La siguiente consulta muestra los detalles de la consulta descargada, incluidos el formato, las filas y el recuento de archivos del comando unload.

```
SELECT query_id,
       file_format,
       start_time,
       duration,
       unloaded_rows,
       unloaded_files_count
FROM sys_unload_history
ORDER BY query_id,
file_format limit 100;
```

Resultados de ejemplo.

```
 query_id | file_format |         start_time         | duration | unloaded_rows |
 unloaded_files_count
----------+-------------+----------------------------+----------+---------------
+----------------------
   527067 | Text        | 2022-02-09 05:18:35.844452 |  5932478 |            10 |
             1
```

## SYS_USERLOG

Registra detalles de los siguientes cambios realizados en un usuario de una base de datos:

- Creación de usuario
- Eliminación de usuario
- Modificación de usuario (cambio de nombre)
- Modificación de usuario (modificación de propiedades)

Puede consultar esta vista para obtener información sobre los grupos de trabajo sin servidor y los clústeres aprovisionados.

SYS_USERLOG solo es visible para los superusuarios. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | El identificador del usuario que envió la descarga. |
| user_name | character (50) | Nombre del usuario afectado por el cambio. |
| original_user_name | character (50) | Nombre de usuario original en una acción de cambio de nombre. Este campo está vacío para el resto de las acciones. |
| acción | character (10) | Acción que sucedió. Los valores válidos son alter, create, drop y rename. |
| has_create_db_privs | entero | Si es true (un valor de 1), el usuario ha creado permisos de base de datos. |
| is_superuser | entero | Si es true (un valor de 1), el usuario puede actualizar los catálogos del sistema. |
| has_update_catalog_privs | entero | Si es true (un valor de 1), el usuario puede actualizar los catálogos del sistema. |
| password_expiration | marca de tiempo | Fecha de vencimiento de la contraseña. |
| session_id | entero | El ID del proceso. |
| transaction_id | bigint | El ID de la transacción. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| record_time | marca de tiempo | Hora en UTC del momento en que comenzó la consulta. |

### Consultas de ejemplo

En el siguiente ejemplo, se realizan cuatro acciones del usuario y, luego, se consulta la vista SYS_USERLOG.

```
CREATE USER userlog1 password 'Userlog1';
ALTER USER userlog1 createdb createuser;
ALTER USER userlog1 rename  to userlog2;
DROP user userlog2;

SELECT user_id, user_name, original_user_name, action, has_create_db_privs,
 is_superuser from SYS_USERLOG order by record_time desc;
```

```
user_id |  user_name | original_user_name |  action | has_create_db_privs |
 is_superuser
--------+------------+--------------------+---------+---------------------+------------
    108 |  userlog2  |                    | drop    |                   1 |   1
    108 |  userlog2  |     userlog1       | rename  |                   1 |   1
    108 |  userlog1  |                    | alter   |                   1 |   1
    108 |  userlog1  |                    | create  |                   0 |   0
 (4 rows)
```

## SYS_VACUUM_HISTORY

Utilice SYS_VACUUM_HISTORY para visualizar los detalles de las consultas de los usuarios. Para obtener información sobre el comando VACUUM, consulte VACUUM.

SYS_VACUUM_HISTORY es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| user_id | entero | ID del usuario que inició la consulta. |
| transaction_id | long | ID de la transacción para la instrucción VACUUM. |
| query_id | long | Identificador de consulta de la instrucción VACUUM. Puede unir esta tabla a la vista SYS_QUERY_DETAIL para ver las instrucciones SQL individuales que se ejecutan para una transacción VACUUM determinada. Si limpia toda la base de datos, cada tabla se limpia en una transacción independiente. Para las operaciones de VACUUM automatizadas, este valor es nulo. |
| database_name | texto | El nombre de la base de datos. |
| schema_name | texto | El nombre del esquema. |
| table_name | texto | El nombre de la tabla. |
| table_id | entero | ID de la tabla. |
| vacuum_type | character | El tipo de la operación VACUUM. Los valores posibles son los siguientes:<br>• **`Delete`** |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | • **`Sort`**<br>• **`Reindex`**<br>• **`Recluster`**<br>• **`Full`**<br><br>Para obtener más informaci ón sobre los tipos de limpieza, consulte VACUUM. |
| is_automatic | booleano | `true` si la operación es una limpieza automática. De lo contrario, `false`. |
| status | character | Descripción de la actividad actual que se realiza como parte de la operación de limpieza:<br>• Initialize<br>• Ordenar<br>• Merge<br>• Eliminar<br>• Seleccionar<br>• Con error<br>• Completado<br>• Omitido<br>• Creación de un orden INTERLEAVED SORTKEY |
| start_time | marca de tiempo | Hora de inicio de la operación de limpieza. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| end_time | marca de tiempo | Hora de finalización de la operación de limpieza. Si la operación está en curso, este campo está en blanco. |
| record_time | marca de tiempo | Hora en que se registró la operación de limpieza en SYS_VACUUM_HISTORY. |
| duration | entero | Número de microsegundos entre el inicio y el final de la operación de limpieza. Si la operación de limpieza está en curso, este campo está en blanco. |
| rows_before_vacuum | bigint | La cantidad real de filas de la tabla más las filas eliminadas que siguen almacenadas en el disco (a la espera de ser limpiadas). |
| size_before_vacuum | entero | Tamaño de la tabla antes de iniciar la operación de limpieza, en MB. |
| reclaimable_rows | bigint | Número de filas que la operación de limpieza estima que recuperará antes de comenzar. |
| reclaimed_rows | bigint | Número de filas que ha recuperado la operación de limpieza. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| reclaimed_blocks | bigint | Número de bloques que ha recuperado la operación de limpieza. |
| sortedrows_before_vacuum | entero | Número de filas ordenadas en la tabla antes de iniciarse la operación de limpieza. |
| sortedrows_after_vacuum | entero | Número adicional de filas ordenadas en la tabla tras finalizar la operación de limpieza. No se incluyen las filas contabilizadas en `sortedrows_before_vacuum` . |

# Asignación de la vista de sistema para migrar a las vistas de supervisión SYS

Al migrar su clúster de Amazon Redshift aprovisionado a Amazon Redshift sin servidor, es posible que sus consultas de monitorización o diagnóstico hagan referencia a vistas del sistema que solo están disponibles en clústeres aprovisionados. Puede actualizar sus consultas para utilizar las vistas de monitorización SYS. Esta página proporciona asignaciones de vistas aprovisionadas solo a SYS para que pueda consultarlas cuando actualice las consultas.

Temas

- SYS_QUERY_HISTORY
- SYS_QUERY_DETAIL
- SYS_RESTORE_LOG
- SYS_RESTORE_STATE
- SYS_TRANSACTION_HISTORY
- SYS_QUERY_TEXT
- SYS_CONNECTION_LOG

- SYS_SESSION_HISTORY
- SYS_LOAD_DETAIL
- SYS_LOAD_HISTORY
- SYS_LOAD_ERROR_DETAIL
- SYS_UNLOAD_HISTORY
- SYS_UNLOAD_DETAIL
- SYS_COPY_REPLACEMENTS
- SYS_DATASHARE_USAGE_CONSUMER
- SYS_DATASHARE_USAGE_PRODUCER
- SYS_DATASHARE_CROSS_REGION_USAGE
- SYS_DATASHARE_CHANGE_LOG
- SYS_EXTERNAL_QUERY_DETAIL
- SYS_EXTERNAL_QUERY_ERROR
- SYS_VACUUM_HISTORY
- SYS_ANALYZE_HISTORY
- SYS_ANALYZE_COMPRESSION_HISTORY
- SYS_MV_REFRESH_HISTORY
- SYS_MV_STATE
- SYS_PROCEDURE_CALL
- SYS_PROCEDURE_MESSAGES
- SYS_UDF_LOG
- SYS_USERLOG
- SYS_SCHEMA_QUOTA_VIOLATIONS
- SYS_SPATIAL_SIMPLIFY

## SYS_QUERY_HISTORY

Algunas o todas las columnas de las tablas siguientes también están definidas en SYS_QUERY_HISTORY.

- STL_DDLTEXT

- STL_ERROR
- STL_QUERY
- STL_UTILITYTEXT
- STL_WLM_QUERY
- STV_INFLIGHT
- STV_RECENTS
- STV_WLM_QUERY_STATE
- SVL_COMPILE
- SVL_MULTI_STATEMENT_VIOLATIONS
- SVL_QLOG
- SVL_QUERY_QUEUE_INFO
- SVL_STATEMENTTEXT
- SVL_TERMINATE

## SYS_QUERY_DETAIL

Algunas o todas las columnas de las tablas siguientes también están definidas en SYS_QUERY_DETAIL.

- STL_AGGR
- STL_ALERT_EVENT_LOG
- STL_BCAST
- STL_DELETE
- STL_DIST
- STL_EXPLAIN
- STL_HASH
- STL_HASHJOIN
- STL_INSERT
- STL_LIMIT
- STL_MERGE
- STL_MERGEJOIN

- STL_NESTLOOP
- STL_PARSE
- STL_PLAN_INFO
- STL_PROJECT
- STL_QUERY_METRICS
- STL_RETURN
- STL_SAVE
- STL_SCAN
- STL_SORT
- STL_STREAM_SEGS
- STL_UNIQUE
- STL_WINDOW
- STV_EXEC_STATE
- STV_QUERY_METRICS
- SVCS_QUERY_SUMMARY
- SVL_QUERY_METRICS
- SVL_QUERY_METRICS_SUMMARY
- SVL_QUERY_REPORT
- SVL_QUERY_SUMMARY
- SVV_QUERY_STATE

## SYS_RESTORE_LOG

Algunas o todas las columnas de la tabla siguiente también están definidas en SYS_RESTORE_LOG.

- SVL_RESTORE_ALTER_TABLE_PROGRESS

## SYS_RESTORE_STATE

Algunas o todas las columnas de la tabla siguiente también están definidas en SYS_RESTORE_STATE.

- STV_XRESTORE_ALTER_QUEUE_STATE

## SYS_TRANSACTION_HISTORY

Algunas o todas las columnas de las tablas siguientes también están definidas en SYS_TRANSACTION_HISTORY.

- STL_COMMIT_STATS
- STL_TR_CONFLICT
- STL_UNDONE

## SYS_QUERY_TEXT

Algunas o todas las columnas de la tabla siguiente también están definidas en SYS_QUERY_TEXT.

- STL_QUERYTEXT

## SYS_CONNECTION_LOG

Algunas o todas las columnas de la tabla siguiente también están definidas en SYS_CONNECTION_LOG.

- STL_CONNECTION_LOG

## SYS_SESSION_HISTORY

Algunas o todas las columnas de las tablas siguientes también están definidas en SYS_SESSION_HISTORY.

- STL_SESSIONS
- STL_RESTARTED_SESSIONS
- STV_SESSIONS

## SYS_LOAD_DETAIL

Algunas o todas las columnas de la tabla siguiente también están definidas en SYS_LOAD_DETAIL.

- STL_LOAD_COMMITS

## SYS_LOAD_HISTORY

Algunas o todas las columnas de la tabla siguiente también están definidas en SYS_LOAD_HISTORY.

- STL_LOAD_COMMITS

## SYS_LOAD_ERROR_DETAIL

Algunas o todas las columnas de las tablas siguientes también están definidas en SYS_LOAD_ERROR_DETAIL.

- STL_LOADERROR_DETAIL
- STL_LOAD_ERRORS

## SYS_UNLOAD_HISTORY

Algunas o todas las columnas de la tabla siguiente también están definidas en SYS_UNLOAD_HISTORY.

- STL_UNLOAD_LOG

## SYS_UNLOAD_DETAIL

Algunas o todas las columnas de la tabla siguiente también están definidas en SYS_UNLOAD_DETAIL.

- STL_UNLOAD_LOG

## SYS_COPY_REPLACEMENTS

Algunas o todas las columnas de la tabla siguiente también están definidas en SYS_COPY_REPLACEMENTS.

- STL_REPLACEMENTS

## SYS_DATASHARE_USAGE_CONSUMER

Algunas o todas las columnas de la tabla siguiente también están definidas en SYS_DATASHARE_USAGE_CONSUMER.

- SVL_DATASHARE_USAGE_CONSUMER

## SYS_DATASHARE_USAGE_PRODUCER

Algunas o todas las columnas de la tabla siguiente también están definidas en SYS_DATASHARE_USAGE_PRODUCER.

- SVL_DATASHARE_USAGE_PRODUCER

## SYS_DATASHARE_CROSS_REGION_USAGE

Algunas o todas las columnas de la tabla siguiente también están definidas en SYS_DATASHARE_CROSS_REGION_USAGE.

- SVL_DATASHARE_CROSS_REGION_USAGE

## SYS_DATASHARE_CHANGE_LOG

Algunas o todas las columnas de la tabla siguiente también están definidas en SYS_DATASHARE_CHANGE_LOG.

- SVL_DATASHARE_CHANGE_LOG

## SYS_EXTERNAL_QUERY_DETAIL

Algunas o todas las columnas de las tablas siguientes también están definidas en SYS_EXTERNAL_QUERY_DETAIL.

- SVL_FEDERATED_QUERY
- SVL_S3LIST
- SVL_S3QUERY
- SVL_S3QUERY_SUMMARY

## SYS_EXTERNAL_QUERY_ERROR

Algunas o todas las columnas de las tablas siguientes también están definidas en SYS_EXTERNAL_QUERY_ERROR.

- SVL_SPECTRUM_SCAN_ERROR

## SYS_VACUUM_HISTORY

Algunas o todas las columnas de las tablas siguientes también están definidas en SYS_VACUUM_HISTORY.

- STL_VACUUM
- SVL_VACUUM_PERCENTAGE
- SVV_VACUUM_PROGRESS
- SVV_VACUUM_SUMMARY

## SYS_ANALYZE_HISTORY

Algunas o todas las columnas de las tablas siguientes también están definidas en SYS_ANALYZE_HISTORY.

- STL_ANALYZE

## SYS_ANALYZE_COMPRESSION_HISTORY

Algunas o todas las columnas de las tablas siguientes también están definidas en SYS_ANALYZE_COMPRESSION_HISTORY.

- STL_ANALYZE_COMPRESSION

## SYS_MV_REFRESH_HISTORY

Algunas o todas las columnas de las tablas siguientes también están definidas en SYS_MV_REFRESH_HISTORY.

- SVL_MV_REFRESH_STATUS

## SYS_MV_STATE

Algunas o todas las columnas de las tablas siguientes también están definidas en SYS_MV_STATE.

- STL_MV_STATE

## SYS_PROCEDURE_CALL

Algunas o todas las columnas de las tablas siguientes también están definidas en SYS_PROCEDURE_CALL.

- SVL_STORED_PROC_CALL

## SYS_PROCEDURE_MESSAGES

Algunas o todas las columnas de las tablas siguientes también están definidas en SYS_PROCEDURE_MESSAGES.

- SVL_STORED_PROC_MESSAGES

## SYS_UDF_LOG

Algunas o todas las columnas de las tablas siguientes también están definidas en SYS_UDF_LOG.

- SVL_UDF_LOG

## SYS_USERLOG

Algunas o todas las columnas de las tablas siguientes también están definidas en SYS_USERLOG.

- STL_USERLOG

## SYS_SCHEMA_QUOTA_VIOLATIONS

Algunas o todas las columnas de las tablas siguientes también están definidas en SYS_SCHEMA_QUOTA_VIOLATIONS.

- STL_SCHEMA_QUOTA_VIOLATIONS

## SYS_SPATIAL_SIMPLIFY

Algunas o todas las columnas de las tablas siguientes también están definidas en SYS_SPATIAL_SIMPLIFY.

- SVL_SPATIAL_SIMPLIFY

# Monitoreo del sistema (solo aprovisionado)

Se pueden consultar las siguientes tablas y vistas del sistema en los clústeres aprovisionados. Las vistas y tablas STL y STV incluyen un subconjunto de datos que se encuentra en distintas tablas de sistema. Estas vistas y tablas proporcionan un acceso más rápido y fácil a los datos más consultados que se encuentran en esas tablas.

Las vistas SVCS proporcionan información sobre las consultas de los clústeres principal y de escalado de simultaneidad. Las vistas SVL proporcionan información solo para las consultas que se ejecutan en el clúster principal, con la excepción de SVL_STATEMENTTEXT. SVL_STATEMENTTEXT puede contener información para consultas ejecutadas en clústeres de escalado simultáneo, así como en el clúster principal.

Temas

- Vistas STL para registro
- Tablas STV para datos de instantáneas
- Vistas de SVCS para clústeres de escalado principales y de simultaneidad
- Vistas de SVL para el clúster principal

## Vistas STL para registro

Las vistas de sistema STL se generan a partir de archivos de registro de Amazon Redshift para proporcionar un historial del sistema.

Estos archivos residen en cada uno de los nodos del clúster en el data warehouse. Las vistas STL toman la información de los registros y les dan el formato de vistas, que pueden utilizar los administradores del sistema.

Retención de registros: las vistas del sistema STL retienen siete días del historial de registros. La retención de registros está garantizada para todos los tamaños y tipos de nodos de clústeres y no

se ve afectada por los cambios en la carga de trabajo del clúster. La retención de registros tampoco se ve afectada por el estado del clúster, como cuando el clúster está en pausa. Tiene menos de siete días de historial de registros solo en el caso de que el clúster sea nuevo. No necesita ninguna acción para retener registros, pero deberá copiar periódicamente los datos de registros en otras tablas o descargarlos en Amazon S3 para conservar los datos de registros con más de siete días de antigüedad.

Temas


- STL_AGGR
- STL_ALERT_EVENT_LOG
- STL_ANALYZE
- STL_ANALYZE_COMPRESSION
- STL_BCAST
- STL_COMMIT_STATS
- STL_CONNECTION_LOG
- STL_DDLTEXT
- STL_DELETE
- STL_DISK_FULL_DIAG
- STL_DIST
- STL_ERROR
- STL_EXPLAIN
- STL_FILE_SCAN
- STL_HASH
- STL_HASHJOIN
- STL_INSERT
- STL_LIMIT
- STL_LOAD_COMMITS
- STL_LOAD_ERRORS
- STL_LOADERROR_DETAIL
- STL_MERGE
- STL_MERGEJOIN
- STL_MV_STATE

- STL_NESTLOOP
- STL_PARSE
- STL_PLAN_INFO
- STL_PROJECT
- STL_QUERY
- STL_QUERY_METRICS
- STL_QUERYTEXT
- STL_REPLACEMENTS
- STL_RESTARTED_SESSIONS
- STL_RETURN
- STL_S3CLIENT
- STL_S3CLIENT_ERROR
- STL_SAVE
- STL_SCAN
- STL_SCHEMA_QUOTA_VIOLATIONS
- STL_SESSIONS
- STL_SORT
- STL_SSHCLIENT_ERROR
- STL_STREAM_SEGS
- STL_TR_CONFLICT
- STL_UNDONE
- STL_UNIQUE
- STL_UNLOAD_LOG
- STL_USAGE_CONTROL
- STL_USERLOG
- STL_UTILITYTEXT
- STL_VACUUM
- STL_WINDOW
- STL_WLM_ERROR
- STL_WLM_RULE_ACTION

- STL_WLM_QUERY

## STL_AGGR

Analiza los pasos de ejecución de agregación de las consultas. Estos pasos tienen lugar durante la ejecución de las funciones de agregación y de las cláusulas GROUP BY.

STL_AGGR es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_AGGR solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| slice | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |
| segment | entero | Número que identifica el segmento de la consulta. |
| paso | entero | Paso de la consulta que se ejecutó. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| starttime | marca de tiempo | Hora en UTC en que comenzó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| endtime | marca de tiempo | Hora en UTC en la que finalizó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| tasknum | entero | Es el número del proceso de tarea de la consulta que se asignó para ejecutar el paso. |
| rows | bigint | Es la cantidad total de filas que se procesaron. |
| bytes | bigint | Tamaño, en bytes, de todas las filas de resultado del paso. |
| slots | entero | Cantidad de buckets hash. |
| occupied | entero | Cantidad de slots que tienen registros. |
| maxlength | entero | Tamaño del slot más grande. |
| tbl | entero | ID de la tabla. |
| is_diskbased | character (1) | Si es true (t) (verdadero), significa que la consulta se ejecutó como una operación basada en el disco. Si es false (f) (falso), implica que la consulta se ejecutó en la memoria. |
| workmem | bigint | Cantidad de bytes de memoria funcional asignada a este paso. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| type | character (6) | El tipo de paso. Los valores válidos son:<br>• HASHED. Indica que el paso utilizó una agregación desordenada agrupada.<br>• PLAIN. Indica que el paso utilizó una agregación escalar no agrupada.<br>• SORTED. Indica que el paso utilizó una agregación ordenada agrupada. |
| resizes | entero | Esta información es solo para uso interno. |
| flushable | entero | Esta información es solo para uso interno. |

Consultas de ejemplo

Devuelve información acerca de los pasos de ejecución de agregación para SLICE 1 y TBL 239.

```
select query, segment, bytes, slots, occupied, maxlength, is_diskbased, workmem, type
from stl_aggr where slice=1 and tbl=239
order by rows
limit 10;
```

```
 query | segment | bytes |  slots  | occupied | maxlength | is_diskbased |  workmem  |
 type
-------+---------+-------+---------+----------+-----------+--------------+-----------
+--------
   562 |       1 |     0 | 4194304 |        0 |         0 | f            | 383385600 |
 HASHED
   616 |       1 |     0 | 4194304 |        0 |         0 | f            | 383385600 |
 HASHED
   546 |       1 |     0 | 4194304 |        0 |         0 | f            | 383385600 |
 HASHED
   547 |       0 |     8 |       0 |        0 |         0 | f            |         0 |
 PLAIN
   685 |       1 |    32 | 4194304 |        1 |         0 | f            | 383385600 |
 HASHED
```

```
  652 |       0 |     8 |       0 |       0 |         0 | f             |         0 |
 PLAIN
  680 |       0 |     8 |       0 |       0 |         0 | f             |         0 |
 PLAIN
  658 |       0 |     8 |       0 |       0 |         0 | f             |         0 |
 PLAIN
  686 |       0 |     8 |       0 |       0 |         0 | f             |         0 |
 PLAIN
  695 |       1 |    32 | 4194304 |       1 |         0 | f             | 383385600 |
 HASHED
(10 rows)
```

## STL_ALERT_EVENT_LOG

Registra una alerta cuando el optimizador de consultas identifica condiciones que podrían indicar problemas de rendimiento. Utilice la vista STL_ALERT_EVENT_LOG para identificar oportunidades de mejora del rendimiento de las consultas.

Una consulta consta de distintos segmentos y cada segmento consta de uno o más pasos. Para obtener más información, consulte Procesamiento de consultas.

STL_ALERT_EVENT_LOG es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_ALERT_EVENT_LOG solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| slice | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |
| segment | entero | Número que identifica el segmento de la consulta. |
| paso | entero | Paso de la consulta que se ejecutó. |
| pid | entero | ID de proceso asociado a la instrucción y al sector. La misma consulta podría tener varios PID si se ejecutara en varios sectores. |
| xid | bigint | ID de la transacción asociado a la instrucción. |
| event | character (1024) | Descripción del evento de alerta. |
| solution | character (1024) | Solución recomendada. |
| event_time | marca de tiempo | Hora en UTC en que comenzó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |

Notas de uso

Puede utilizar STL_ALERT_EVENT_LOG para identificar posibles problemas en sus consultas; luego, siga las prácticas descritas en Ajuste del rendimiento de las consultas para optimizar el diseño de su base de datos y vuelva a escribir sus consultas. STL_ALERT_EVENT_LOG registra las siguientes alertas:

- Faltan estadísticas

  Faltan estadísticas. Ejecute la función ANALYZE después de la carga de datos o de actualizaciones importantes y utilice la función STATUPDATE con las operaciones COPY. Para obtener más información, consulte Prácticas recomendadas de Amazon Redshift para el diseño de consultas.

- Bucle anidado

  Un bucle anidado es, por lo general, un producto cartesiano. Evalúe su consulta para asegurarse de que todas las tablas que emplee estén combinadas de forma eficaz.

- Filtro muy selectivo

  La relación de filas devueltas respecto a las filas examinadas es inferior al 0,05. Las filas examinadas es el valor de `rows_pre_user_filter` y las filas devueltas es el valor de las filas de la vista de sistema STL_SCAN. Indica que la consulta está examinando una cantidad excepcionalmente grande de filas para determinar el conjunto de resultados. Esto puede deberse a que faltan claves de ordenación, o que estas son incorrectas. Para obtener más información, consulte Claves de clasificación.

- Filas fantasma excesivas

  Un examen omitió una cantidad considerablemente grande de filas que están marcadas como eliminadas pero no limpiadas o filas que fueron insertadas pero no confirmadas. Para obtener más información, consulte Limpieza de tablas.

- Distribución grande

  Más de 1 000 000 filas fueron redistribuidas para una combinación hash o una agregación. Para obtener más información, consulte Distribución de datos para la optimización de consultas.

- Difusión grande

  Más de 1 000 000 filas fueron difundidas para una combinación hash. Para obtener más información, consulte Distribución de datos para la optimización de consultas.

- Ejecución en serie

  En el plan de consulta se indicó un estilo de redistribución DS_DIST_ALL_INNER, lo que exige una ejecución en serie porque toda la tabla interna fue redistribuida a un nodo único. Para obtener más información, consulte Distribución de datos para la optimización de consultas.

Consultas de ejemplo

La siguiente consulta muestra eventos de alerta para cuatro consultas.

```
SELECT query, substring(event,0,25) as event,
substring(solution,0,25) as solution,
trim(event_time) as event_time from stl_alert_event_log order by query;

 query |             event             |          solution            |     event_time

-------+-------------------------------+-------------------------------
+---------------------
  6567 | Missing query planner statist | Run the ANALYZE command       | 2014-01-03
 18:20:58
  7450 | Scanned a large number of del | Run the VACUUM command to rec| 2014-01-03
 21:19:31
  8406 | Nested Loop Join in the query | Review the join predicates to| 2014-01-04
 00:34:22
 29512 | Very selective query filter:r | Review the choice of sort key| 2014-01-06
 22:00:00

(4 rows)
```

## STL_ANALYZE

Registra detalles para las operaciones ANALYZE.

SYS_ANALYZE solo está visible para los superusuarios. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_ANALYZE_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| xid | long | El ID de la transacción. |
| database | char (30) | El nombre de la base de datos. |
| table_id | entero | El ID de la tabla. |
| status | char (15) | El resultado del comando analyze. Los posibles valores son `Full`, `Skipped` y `PredicateColumn` . |
| rows | double | La cantidad total de filas en la tabla. |
| modified_rows | double | La cantidad total de filas que se modificaron desde la última operación ANALYZE. |
| threshold_percent | entero | El valor del parámetro `analyze_threshold_percent` . |
| is_auto | char(1) | El valor es true (`t`) si la operación incluía una operación analyze de Amazon Redshift de manera predeterminada. El valor es false (`f`) si el comando ANALYZE se ejecutaba explícitamente. |
| starttime | marca de tiempo | La hora en UTC en que comenzó a ejecutarse la operación analyze. |
| endtime | marca de tiempo | La hora en UTC en que terminó de ejecutarse la operación analyze. |
| prevtime | marca de tiempo | La hora en UTC en que se analizó previamente la tabla. |
| num_predicate_cols | entero | La cantidad actual de columnas de predicados en la tabla. |
| num_new_predicate_cols | entero | La cantidad de columnas nuevas de predicado desde la operación analyze anterior. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| is_background | character (1) | El valor es true (t) si una operación analyze automática ejecutaba el análisis. De lo contrario, el valor es false (f). |
| auto_analyze_phase | character (100) | Se reserva para uso interno. |
| schema_name | char(128) | Nombre del esquema para la tabla. |
| table_name | char (136) | El nombre de la tabla. |

Consultas de ejemplo

En el siguiente ejemplo, se combina STV_TBL_PERM para mostrar el nombre de la tabla y los detalles de ejecución.

```
select distinct a.xid, trim(t.name) as name, a.status, a.rows, a.modified_rows,
 a.starttime, a.endtime
from stl_analyze a
join stv_tbl_perm t  on t.id=a.table_id
where name = 'users'
order by starttime;

xid    | name  | status          | rows  | modified_rows | starttime           |
 endtime
-------+-------+-----------------+-------+---------------+---------------------
+---------------------
  1582 | users | Full            | 49990 |         49990 | 2016-09-22 22:02:23 |
 2016-09-22 22:02:28
244287 | users | Full            | 24992 |         74988 | 2016-10-04 22:50:58 |
 2016-10-04 22:51:01
244712 | users | Full            | 49984 |         24992 | 2016-10-04 22:56:07 |
 2016-10-04 22:56:07
245071 | users | Skipped         | 49984 |             0 | 2016-10-04 22:58:17 |
 2016-10-04 22:58:17
245439 | users | Skipped         | 49984 |          1982 | 2016-10-04 23:00:13 |
 2016-10-04 23:00:13
```

```
(5 rows)
```

## STL_ANALYZE_COMPRESSION

Registra detalles para las operaciones de análisis de compresión durante los comandos COPY o ANALYZE COMPRESSION.

SYS_ANALYZE_COMPRESSION es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_ANALYZE_COMPRESSION_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| start_time | marca de tiempo | El momento en que la operación de análisis de compresión se inició. |
| xid | bigint | La ID de transacción de la operación de análisis de compresión. |
| tbl | entero | La ID de tabla de la tabla que se analizó. |
| tablename | character (128) | El nombre de la tabala que se analizó. |
| col | entero | El índice de la columna de la tabla que se analizó para determinar la codificación de la compresión. |
| old_encoding | character (15) | El tipo de codificación antes del análisis de compresión. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| new_encoding | character (15) | El tipo de codificación después del análisis de compresión. |
| mode | Carácter(14) | Los valores posibles son:<br>PRESET<br>Especifica que `new_encoding` está determinado por el comando COPY de Amazon Redshift según el tipo de datos de la columna. No se realiza un muestreo de los datos.<br>ON<br>Especifica que `new_encoding` está determinado por el comando COPY de Amazon Redshift según un análisis de datos de muestra.<br>ANALYZE ONLY<br>Especifica que `new_encoding` está determinado por el comando ANALYZE COMPRESSION de Amazon Redshift según un análisis de datos de muestra. No obstante, el tipo de codificación de la columna analizada no ha cambiado. |
| best_compression_encoding | character (15) | El tipo de codificación que proporciona la mejor relación de compresión. |
| recommended_bytes | character (15) | Los bytes utilizados al adoptar la nueva codificación. |
| best_compression_bytes | character (15) | Los bytes utilizados al adoptar la nueva mejor codificación de compresión. |
| ndv | bigint | El número de valores distintos en las filas muestreadas. |

## Consultas de ejemplo

El siguiente ejemplo inspecciona los detalles del análisis de compresión en la tabla `lineitem` por el último comando COPY ejecutado en la misma sesión.

```
select xid, tbl, btrim(tablename) as tablename, col, old_encoding, new_encoding,
 best_compression_encoding, mode
from stl_analyze_compression
where xid = (select xid from stl_query where query = pg_last_copy_id()) order by col;

 xid  |  tbl   | tablename | col |  old_encoding   |  new_encoding   |
 best_compression_encoding |      mode
------+--------+-----------+-----+-----------------+-----------------
+---------------------------+----------------
 5308 | 158961 | $lineitem |   0 | mostly32        | az64            | delta
          | ON
 5308 | 158961 | $lineitem |   1 | mostly32        | az64            | az64
          | ON
 5308 | 158961 | $lineitem |   2 | lzo             | az64            | az64
          | ON
 5308 | 158961 | $lineitem |   3 | delta           | az64            | az64
          | ON
 5308 | 158961 | $lineitem |   4 | bytedict        | az64            | bytedict
          | ON
 5308 | 158961 | $lineitem |   5 | mostly32        | az64            | az64
          | ON
 5308 | 158961 | $lineitem |   6 | delta           | az64            | az64
          | ON
 5308 | 158961 | $lineitem |   7 | delta           | az64            | az64
          | ON
 5308 | 158961 | $lineitem |   8 | lzo             | lzo             | lzo
          | ON
 5308 | 158961 | $lineitem |   9 | runlength       | runlength       | runlength
          | ON
 5308 | 158961 | $lineitem |  10 | delta           | az64            | az64
          | ON
 5308 | 158961 | $lineitem |  11 | delta           | az64            | az64
          | ON
 5308 | 158961 | $lineitem |  12 | delta           | az64            | az64
          | ON
 5308 | 158961 | $lineitem |  13 | bytedict        | bytedict        | bytedict
          | ON
 5308 | 158961 | $lineitem |  14 | bytedict        | bytedict        | bytedict
          | ON
```

```
 5308 | 158961 | $lineitem |  15 | text255         | text255         | text255
          | ON
(16 rows)
```

## STL_BCAST

Registra información acerca de la actividad de red durante la ejecución de los pasos de la consulta que difunden datos. Los números de filas, bytes y paquetes que se envían por la red durante un paso determinado en un sector determinado capturan el tráfico de red. La duración del paso es la diferencia entre la hora de inicio y de finalización del registro.

Para identificar los pasos de difusión en una consulta, busque las etiquetas bcast en la vista SVL_QUERY_SUMMARY o ejecute el comando EXPLAIN y, luego, busque atributos del paso que incluyan la etiqueta bcast.

STL_BCAST es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_BCAST solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| slice | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |
| segment | entero | Número que identifica el segmento de la consulta. |
| paso | entero | Paso de la consulta que se ejecutó. |
| starttime | marca de tiempo | Hora en UTC en que comenzó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| endtime | marca de tiempo | Hora en UTC en la que finalizó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| tasknum | entero | Es el número del proceso de tarea de la consulta que se asignó para ejecutar el paso. |
| rows | bigint | Es la cantidad total de filas que se procesaron. |
| bytes | bigint | Tamaño, en bytes, de todas las filas de resultado del paso. |
| packets | entero | Cantidad total de paquetes enviados por la red. |

Consultas de ejemplo

El siguiente ejemplo devuelve información de difusión para las consultas donde hay uno o más paquetes, y la diferencia entre el inicio y la finalización de la consulta fue de un segundo o más.

```
select query, slice, step, rows, bytes, packets, datediff(seconds, starttime, endtime)
from stl_bcast
where packets>0 and datediff(seconds, starttime, endtime)>0;
```

```
 query | slice | step | rows | bytes | packets | date_diff
```

```
-------+-------+------+------+-------+---------+-----------
   453 |     2 |    5 |    1 |   264 |       1 |         1
   798 |     2 |    5 |    1 |   264 |       1 |         1
  1408 |     2 |    5 |    1 |   264 |       1 |         1
  2993 |     0 |    5 |    1 |   264 |       1 |         1
  5045 |     3 |    5 |    1 |   264 |       1 |         1
  8073 |     3 |    5 |    1 |   264 |       1 |         1
  8163 |     3 |    5 |    1 |   264 |       1 |         1
  9212 |     1 |    5 |    1 |   264 |       1 |         1
  9873 |     1 |    5 |    1 |   264 |       1 |         1
(9 rows)
```

## STL_COMMIT_STATS

Proporciona métricas relacionadas con el rendimiento de las confirmaciones, incluido el tiempo de distintas fases de confirmación y la cantidad de bloques confirmados. Consulte STL_COMMIT_STATS para determinar qué parte de una transacción se utilizó en la confirmación y cuántas colas hay en curso.

STL_COMMIT_STATS solo es visible para los superusuarios. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_TRANSACTION_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| xid | bigint | Se está confirmando el id de la transacción. |
| node | integer | Número de nodo. -1 es el nodo principal. |
| startqueue | timestamp | Inicio de la cola para confirmación. |
| startwork | timestamp | Inicio de la confirmación. |
| endflush | timestamp | Finalización de la fase de vaciado de bloques sucios. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| endstage | timestamp | Finalización de la fase provisional de metadatos. |
| endlocal | timestamp | Finalización de la fase de confirmación local. |
| startglobal | timestamp | Inicio de la fase global. |
| endtime | timestamp | Finalización de la confirmación. |
| queuelen | bigint | Cantidad de transacciones que estaban antes de esta transacción en la cola de confirmación. |
| permblocks | bigint | Cantidad de bloques permanentes existentes en el momento de esta confirmación. |
| newblocks | bigint | Cantidad de bloques permanentes nuevos en el momento de esta confirmación. |
| dirtyblocks | bigint | Cantidad de bloques que tienen que escribirse como parte de esta confirmación. |
| headers | bigint | Cantidad de encabezados de bloques que tienen que escribirse como parte de esta confirmación. |
| numxids | integer | El número de transacciones de DML activas. |
| oldestxid | bigint | El XID de la transacción de DML activa más antigua. |
| extwritelatency | bigint | Esta información es solo para uso interno. |
| metadatawritten | int | Esta información es solo para uso interno. |
| tombstonedblocks | bigint | Esta información es solo para uso interno. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| tossedblocks | bigint | Esta información es solo para uso interno. |
| batched_by | bigint | Esta información es solo para uso interno. |

Consulta de ejemplo

```
select node, datediff(ms,startqueue,startwork) as queue_time,
datediff(ms, startwork, endtime) as commit_time, queuelen
from stl_commit_stats
where xid = 2574
order by node;

node | queue_time   | commit_time | queuelen
-----+--------------+-------------+---------
  -1 |            0 |         617 |        0
   0 | 444950725641 |         616 |        0
   1 | 444950725636 |         616 |        0
```

## STL_CONNECTION_LOG

Registra los intentos de autenticación y las conexiones y desconexiones.

STL_CONNECTION_LOG solo es visible para los superusuarios. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_CONNECTION_LOG. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| event | character(50) | Evento de conexión o de autenticación. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| recordtime | marca de tiempo | Hora cuando se produjo el evento. |
| remotehost | character(45) | Nombre o dirección IP del host remoto. |
| remoteport | character (32) | Número del puerto para el host remoto. |
| pid | entero | ID del proceso asociado a la instrucción. |
| dbname | character (50) | Nombre de la base de datos. |
| nombre de usuario | character (50) | Nombre de usuario. |
| authmethod | character (32) | Método de autenticación. |
| duration | entero | Duración de la conexión en microsegundos. |
| sslversion | character (50) | Versión de la capa de conexión segura (SSL). Esta columna está vacía cuando SSL está desactivado. |
| sslcipher | character (128) | Cifrado de la SSL. |
| mtu | entero | Unidad de transmisión máxima (MTU). |
| sslcompression | character(64) | Tipo de compresión de la SSL. |
| sslexpansion | character (64) | Tipo de expansión de la SSL. |
| iamauthguid | character (36) | ID de autenticación de IAM para la solicitud de CloudTrail. |
| application_name | character (250) | Nombre original o actualizado de la aplicación para una sesión. |
| os_version | character(64) | La versión del sistema operativo que se encuentra en el equipo cliente que se conecta a su clúster de Amazon Redshift. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| driver_version | character(64) | La versión del controlador ODBC o JDBC que se conecta a su clúster de Amazon Redshift desde las herramientas de cliente SQL de terceros. |
| plugin_name | character (32) | El nombre del complemento utilizado para conectarse a su clúster de Amazon Redshift. |
| protocol_version | entero | La versión del protocolo interno que utiliza el controlador de Amazon Redshift cuando se establece su conexión con el servidor. Las versiones de protocolo se negocian entre el controlador y el servidor. En la versión se describen las características disponibles. Los valores válidos son:<br>• 0 (BASE_SERVER_PROTOCOL_VERSION)<br>• 1 (EXTENDED_RESULT_METADATA_SERVER_PROTOCOL_VERSION): para guardar un viaje de ida y vuelta por consulta, el servidor envía información adicional de metadatos del conjunto de resultados.<br>• 2 (BINARY_PROTOCOL_VERSION): según el tipo de datos del conjunto de resultados, el servidor envía datos en formato binario.<br>• 3 (EXTENDED2_RESULT_METADATA_SERVER_PROTOCOL_VERSION): el servidor envía información (de intercalación) con distinción de mayúsculas y minúsculas de una columna. |
| sessionid | character (36) | El identificador único global de la sesión actual. El ID de la sesión persiste durante los reinicios de errores de nodo. |
| compression | character(16) | Algoritmo de compresión que se utiliza para la conexión. |

## Consultas de ejemplo

Para ver los detalles de las conexiones abiertas, ejecute la siguiente consulta.

```
select recordtime, username, dbname, remotehost, remoteport
from stl_connection_log
where event = 'initiating session'
and pid not in
(select pid from stl_connection_log
where event = 'disconnecting session')
order by 1 desc;

recordtime          | username    | dbname     | remotehost    | remoteport

--------------------+-------------+------------+---------------+------------
2014-11-06 20:30:06 | rdsdb       | dev        | [local]       |

2014-11-06 20:29:37 | test001     | test       | 10.49.42.138  | 11111

2014-11-05 20:30:29 | rdsdb       | dev        | 10.49.42.138  | 33333

2014-11-05 20:28:35 | rdsdb       | dev        | [local]       |
(4 rows)
```

En el siguiente ejemplo, se refleja un intento fallido de autenticación y una correcta conexión y desconexión.

```
select event, recordtime, remotehost, username
from stl_connection_log order by recordtime;

            event      |         recordtime        |  remotehost   | username

-----------------------+---------------------------+---------------+---------
authentication failure | 2012-10-25 14:41:56.96391 | 10.49.42.138 | john

authenticated          | 2012-10-25 14:42:10.87613 | 10.49.42.138 | john

initiating session     | 2012-10-25 14:42:10.87638 | 10.49.42.138 | john

disconnecting session  | 2012-10-25 14:42:19.95992 | 10.49.42.138 | john

(4 rows)
```

En el siguiente ejemplo, se muestra la versión del controlador ODBC, el sistema operativo del equipo cliente y el complemento utilizado para conectarse al clúster de Amazon Redshift. En este ejemplo, el complemento se utiliza para la autenticación de controladores ODBC estándar mediante un nombre de inicio de sesión y una contraseña.

```
select driver_version, os_version, plugin_name from stl_connection_log;

driver_version                          |  os_version                       |
 plugin_name
----------------------------------------+-----------------------------------
+--------------------
Amazon Redshift ODBC Driver 1.4.15.0001 | Darwin 18.7.0 x86_64              | none
Amazon Redshift ODBC Driver 1.4.15.0001 | Linux 4.15.0-101-generic x86_64   | none
```

En el siguiente ejemplo, se muestra la versión del sistema operativo en el equipo cliente, la versión del controlador y la versión del protocolo.

```
select os_version, driver_version, protocol_version from stl_connection_log;

os_version                      |  driver_version              | protocol_version
--------------------------------+------------------------------+--------------------
Linux 4.15.0-101-generic x86_64 | Redshift JDBC Driver 2.0.0.0 | 2
Linux 4.15.0-101-generic x86_64 | Redshift JDBC Driver 2.0.0.0 | 2
Linux 4.15.0-101-generic x86_64 | Redshift JDBC Driver 2.0.0.0 | 2
```

## STL_DDLTEXT

Captura las siguientes instrucciones DDL que se ejecutaron en el sistema.

Estas instrucciones DDL incluyen las siguientes consultas y objetos:

- CREATE SCHEMA, TABLE, VIEW
- DROP SCHEMA, TABLE, VIEW
- ALTER SCHEMA, TABLE

Consulte también STL_QUERYTEXT, STL_UTILITYTEXT y SVL_STATEMENTTEXT. Estas vistas proporcionan una línea temporal de los comandos SQL que se ejecutan en el sistema; este historial es útil para solucionar problemas y para crear una traza de auditoría de todas las actividades del sistema.

Utilice las columnas STARTTIME y ENDTIME para averiguar qué instrucciones se registraron durante un periodo determinado. Los bloques grandes de texto de SQL se dividen en líneas de 200 caracteres. La columna SEQUENCE identifica los segmentos de texto que pertenecen a una única instrucción.

STL_DDLTEXT es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_QUERY_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| xid | bigint | ID de la transacción asociado a la instrucción. |
| pid | entero | ID del proceso asociado a la instrucción. |
| etiqueta | character(320) | Ya sea el nombre del archivo utilizado para ejecutar la consulta o una etiqueta definida con un comando SET QUERY_GROUP. Si la consulta no se basa en archivos o si no se establece el parámetro QUERY_GROUP, el valor del campo queda en blanco. |
| starttime | marca de tiempo | Hora en UTC en que comenzó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| endtime | marca de tiempo | Hora en UTC en la que finalizó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| sequence | entero | Cuando una única instrucción tiene más de 200 caractere s, se registran filas adicionales para esa instrucción. Secuencia 0 es la primera fila, 1 es la segunda y, así, sucesivamente. |
| texto | character(200) | Texto SQL, en incrementos de 200 caracteres. El campo puede contener caracteres especiales como una barra inversa(\\) y una nueva línea (\n). |

Consultas de ejemplo

La siguiente consulta devuelve registros que incluyen instrucciones DDL ejecutadas previamente.

```
select xid, starttime, sequence, substring(text,1,40) as text
from stl_ddltext order by xid desc, sequence;
```

Aquí se incluye un resultado de ejemplo que muestra cuatro instrucciones CREATE TABLE. Las instrucciones DDL aparecen en la columna `text`, que se trunca para mejorar su legibilidad.

```
 xid  |         starttime          | sequence |                  text
------+----------------------------+----------
+----------------------------------------
 1806 | 2013-10-23 00:11:14.709851 |        0 | CREATE TABLE supplier ( s_suppkey int4
 N
 1806 | 2013-10-23 00:11:14.709851 |        1 |  s_comment varchar(101) NOT NULL )
 1805 | 2013-10-23 00:11:14.496153 |        0 | CREATE TABLE region ( r_regionkey int4
 N
 1804 | 2013-10-23 00:11:14.285986 |        0 | CREATE TABLE partsupp ( ps_partkey int8
 1803 | 2013-10-23 00:11:14.056901 |        0 | CREATE TABLE part ( p_partkey int8 NOT
 N
 1803 | 2013-10-23 00:11:14.056901 |        1 | ner char(10) NOT NULL , p_retailprice
 nu
(6 rows)
```

### Reconstruir SQL almacenados

El siguiente SQL muestra las filas almacenadas en la columna `text` de STL_DDLTEXT. Las filas están ordenadas por `xid` y `sequence`. Si el SQL original tenía más de 200 caracteres en varias filas, STL_DDLTEXT puede contener varias filas por `sequence`.

```
SELECT xid, sequence, LISTAGG(CASE WHEN LEN(RTRIM(text)) = 0 THEN text ELSE RTRIM(text)
 END, '') WITHIN GROUP (ORDER BY sequence) as query_statement
FROM stl_ddltext GROUP BY xid, sequence ORDER BY xid, sequence;
```

```
xid     |  sequence | query_statement
--------+-----------+-----------------
7886671    0          create external schema schema_spectrum_uddh\nfrom data catalog
\ndatabase 'spectrum_db_uddh'\niam_role ''\ncreate external database if not exists;
7886752    0          CREATE EXTERNAL TABLE schema_spectrum_uddh.soccer_league\n(\n
  league_rank smallint,\n  prev_rank   smallint,\n  club_name   varchar(15),\n
 league_name varchar(20),\n  league_off  decimal(6,2),\n  le
7886752    1          ague_def  decimal(6,2),\n  league_spi  decimal(6,2),\n
 league_nspi smallint\n)\nROW FORMAT DELIMITED \n    FIELDS TERMINATED BY ',' \n
 LINES TERMINATED BY '\\n\\l'\nstored as textfile\nLOCATION 's
7886752    2          3://mybucket-spectrum-uddh/'\ntable properties
 ('skip.header.line.count'='1');
...
```

Para reconstruir el SQL almacenado en la columna `text` de STL_DDLTEXT, ejecute la siguiente instrucción SQL. Reúne las instrucciones DDL de uno o varios segmentos en la columna `text`. Antes de ejecutar el SQL reconstruido, sustituya cualquier carácter especial (\n) con una nueva línea en el cliente SQL. Los resultados de la siguiente instrucción SELECT reúnen tres filas secuenciales para reconstruir el SQL, en el campo `query_statement`.

```
SELECT LISTAGG(CASE WHEN LEN(RTRIM(text)) = 0 THEN text ELSE RTRIM(text) END) WITHIN
 GROUP (ORDER BY sequence) as query_statement
FROM stl_ddltext GROUP BY xid, endtime order by xid, endtime;
```

```
query_statement
--------------
create external schema schema_spectrum_uddh\nfrom data catalog\ndatabase
 'spectrum_db_uddh'\niam_role ''\ncreate external database if not exists;
```

```
CREATE EXTERNAL TABLE schema_spectrum_uddh.soccer_league\n(\n  league_rank smallint,
\n  prev_rank   smallint,\n  club_name   varchar(15),\n  league_name varchar(20),\n
 league_off  decimal(6,2),\n  league_def  decimal(6,2),\n  league_spi  decimal(6,2),
\n  league_nspi smallint\n)\nROW FORMAT DELIMITED \n    FIELDS TERMINATED BY ',' \n
  LINES TERMINATED BY '\\n\\l'\nstored as textfile\nLOCATION 's3://mybucket-spectrum-
uddh/'\ntable properties ('skip.header.line.count'='1');
```

## STL_DELETE

Analiza los pasos de ejecución de eliminación de las consultas.

STL_DELETE es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_DELETE solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| slice | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |
| segment | entero | Número que identifica el segmento de la consulta. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| paso | entero | Paso de la consulta que se ejecutó. |
| starttime | marca de tiempo | Hora en UTC en que comenzó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **2009-06-12 11:29:19.131358** . |
| endtime | marca de tiempo | Hora en UTC en la que finalizó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **2009-06-12 11:29:19.131358** . |
| tasknum | entero | Es el número del proceso de tarea de la consulta que se asignó para ejecutar el paso. |
| rows | bigint | Es la cantidad total de filas que se procesaron. |
| tbl | entero | ID de la tabla. |

Consultas de ejemplo

Para crear una fila en STL_DELETE, en el siguiente ejemplo se inserta una fila en la tabla EVENT y, luego, se la elimina.

Primero, se inserta una fila en la tabla EVENT y comprueba que se haya insertado.

```
insert into event(eventid,venueid,catid,dateid,eventname)
values ((select max(eventid)+1 from event),95,9,1857,'Lollapalooza');
```

```
select * from event
where eventname='Lollapalooza'
order by eventid;
```

```
 eventid | venueid | catid | dateid |  eventname   |      starttime
---------+---------+-------+--------+--------------+---------------------
```

```
    4274 |     102 |     9 |   1965 | Lollapalooza | 2008-05-01 19:00:00
    4684 |     114 |     9 |   2105 | Lollapalooza | 2008-10-06 14:00:00
    5673 |     128 |     9 |   1973 | Lollapalooza | 2008-05-01 15:00:00
    5740 |      51 |     9 |   1933 | Lollapalooza | 2008-04-17 15:00:00
    5856 |     119 |     9 |   1831 | Lollapalooza | 2008-01-05 14:00:00
    6040 |     126 |     9 |   2145 | Lollapalooza | 2008-11-15 15:00:00
    7972 |      92 |     9 |   2026 | Lollapalooza | 2008-07-19 19:30:00
    8046 |      65 |     9 |   1840 | Lollapalooza | 2008-01-14 15:00:00
    8518 |      48 |     9 |   1904 | Lollapalooza | 2008-03-19 15:00:00
    8799 |      95 |     9 |   1857 | Lollapalooza |
(10 rows)
```

Ahora, se elimina la fila que agregó en la tabla EVENT y se comprueba que se haya eliminado.

```
delete from event
where eventname='Lollapalooza' and eventid=(select max(eventid) from event);
```

```
select * from event
where eventname='Lollapalooza'
order by eventid;
```

```
 eventid | venueid | catid | dateid |  eventname   |      starttime
---------+---------+-------+--------+--------------+---------------------
    4274 |     102 |     9 |   1965 | Lollapalooza | 2008-05-01 19:00:00
    4684 |     114 |     9 |   2105 | Lollapalooza | 2008-10-06 14:00:00
    5673 |     128 |     9 |   1973 | Lollapalooza | 2008-05-01 15:00:00
    5740 |      51 |     9 |   1933 | Lollapalooza | 2008-04-17 15:00:00
    5856 |     119 |     9 |   1831 | Lollapalooza | 2008-01-05 14:00:00
    6040 |     126 |     9 |   2145 | Lollapalooza | 2008-11-15 15:00:00
    7972 |      92 |     9 |   2026 | Lollapalooza | 2008-07-19 19:30:00
    8046 |      65 |     9 |   1840 | Lollapalooza | 2008-01-14 15:00:00
    8518 |      48 |     9 |   1904 | Lollapalooza | 2008-03-19 15:00:00
(9 rows)
```

Luego, se consulta la tabla stl_delete para ver lo pasos de ejecución de la eliminación. En este ejemplo, la consulta devolvió más de 300 filas, por lo que la siguiente salida está abreviada a título informativo.

```
select query, slice, segment, step, tasknum, rows, tbl from stl_delete order by query;
```

```
 query | slice | segment | step | tasknum | rows |  tbl
-------+-------+---------+------+---------+------+--------
     7 |     0 |       0 |    1 |       0 |    0 | 100000
     7 |     1 |       0 |    1 |       0 |    0 | 100000
     8 |     0 |       0 |    1 |       2 |    0 | 100001
     8 |     1 |       0 |    1 |       2 |    0 | 100001
     9 |     0 |       0 |    1 |       4 |    0 | 100002
     9 |     1 |       0 |    1 |       4 |    0 | 100002
    10 |     0 |       0 |    1 |       6 |    0 | 100003
    10 |     1 |       0 |    1 |       6 |    0 | 100003
    11 |     0 |       0 |    1 |       8 |    0 | 100253
    11 |     1 |       0 |    1 |       8 |    0 | 100253
    12 |     0 |       0 |    1 |       0 |    0 | 100255
    12 |     1 |       0 |    1 |       0 |    0 | 100255
    13 |     0 |       0 |    1 |       2 |    0 | 100257
    13 |     1 |       0 |    1 |       2 |    0 | 100257
    14 |     0 |       0 |    1 |       4 |    0 | 100259
    14 |     1 |       0 |    1 |       4 |    0 | 100259
    ...
```

## STL_DISK_FULL_DIAG

Registra información sobre los errores producidos cuando el disco está lleno.

STL_DISK_FULL_DIAG solo es visible para los superusuarios. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| currenttime | bigint | El día y hora el error se generó en microsegundos desde el 1 de enero del 2000. |
| node_num | bigint | Identificador del disco. |
| query_id | bigint | Identificador de la consulta que causó el error. |

| Nombre de la columna | Tipo de datos: | Descripción | | |
|---|---|---|---|---|
| temp_blocks | bigint | Número de bloques temporales creados por la consulta. | | |

Consultas de ejemplo

El siguiente ejemplo devuelve información detallada acerca de los datos almacenados cuando hay un error de disco lleno.

```
select * from stl_disk_full_diag
```

El siguiente ejemplo convierte la cadena `currenttime` a una marca de hora.

```
select '2000-01-01'::timestamp + (currenttime/1000000.0)* interval '1 second' as
 currenttime,node_num,query_id,temp_blocks from pg_catalog.stl_disk_full_diag;
```

```
        currenttime         | node_num | query_id | temp_blocks
----------------------------+----------+----------+-------------
 2019-05-18 19:19:18.609338 |        0 |   569399 |       70982
 2019-05-18 19:37:44.755548 |        0 |   569580 |       70982
 2019-05-20 13:37:20.566916 |        0 |   597424 |       70869
```

## STL_DIST

Registra información acerca de la actividad de red durante la ejecución de los pasos de la consulta que distribuyen datos. Los números de filas, bytes y paquetes que se envían por la red durante un paso determinado en un sector determinado capturan el tráfico de red. La duración del paso es la diferencia entre la hora de inicio y de finalización del registro.

Para identificar los pasos de distribución en una consulta, busque las etiquetas dist en la vista QUERY_SUMMARY o ejecute el comando EXPLAIN y luego busque atributos del paso que incluyan la etiqueta dist.

STL_DIST es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_DIST solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| slice | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |
| segment | entero | Número que identifica el segmento de la consulta. |
| paso | entero | Paso de la consulta que se ejecutó. |
| starttime | marca de tiempo | Hora en UTC en que comenzó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| endtime | marca de tiempo | Hora en UTC en la que finalizó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | para fracciones de segundo. Por ejemplo: **2009-06-12 11:29:19.131358** . |
| tasknum | entero | Es el número del proceso de tarea de la consulta que se asignó para ejecutar el paso. |
| rows | bigint | Es la cantidad total de filas que se procesaron. |
| bytes | bigint | Tamaño, en bytes, de todas las filas de resultado del paso. |
| packets | entero | Cantidad total de paquetes enviados por la red. |

Consultas de ejemplo

El siguiente ejemplo devuelve información acerca de la distribución de las consultas con uno o más paquetes y con duración mayor que cero.

```
select query, slice, step, rows, bytes, packets,
datediff(seconds, starttime, endtime) as duration
from stl_dist
where packets>0 and datediff(seconds, starttime, endtime)>0
order by query
limit 10;
```

```
 query  | slice | step |  rows  |  bytes  | packets | duration
--------+-------+------+--------+---------+---------+----------
    567 |     1 |    4 |  49990 | 6249564 |     707 |        1
    630 |     0 |    5 |   8798 |  408404 |      46 |        2
    645 |     1 |    4 |   8798 |  408404 |      46 |        1
    651 |     1 |    5 | 192497 | 9226320 |    1039 |        6
    669 |     1 |    4 | 192497 | 9226320 |    1039 |        4
    675 |     1 |    5 |   3766 |  194656 |      22 |        1
    696 |     0 |    4 |   3766 |  194656 |      22 |        1
    705 |     0 |    4 |    930 |   44400 |       5 |        1
 111525 |     0 |    3 |     68 |   17408 |       2 |        1
(9 rows)
```

## STL_ERROR

Registra los errores internos de procesamiento generados por el motor de base de datos de Amazon Redshift. STL_ERROR no registra los errores ni los mensajes de SQL. La información en STL_ERROR es útil para solucionar determinados errores. Un ingeniero de AWS Support podría pedirle que le proporcione esta información como parte del proceso de solución de problemas.

STL_ERROR es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_QUERY_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Para ver una lista de los códigos de error que se pueden generar al cargar datos con el comando Copy, consulte Referencia de error de carga.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| process | character(12) | Proceso que arrojó la excepción. |
| recordtime | marca de tiempo | Hora en la que ocurrió el error. |
| pid | entero | ID de proceso. La tabla STL_QUERY tiene los ID de los procesos y los ID de consulta únicos de las consultas completadas. |
| errcode | entero | Código de error que corresponde a la categoría error. |
| archivo | character (90) | Nombre del archivo de origen donde ocurrió el error. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| linenum | entero | Número de línea en el archivo de origen donde ocurrió el error. |
| context | character (100) | Causa del error. |
| error | character (512) | Mensaje de error. |

Consultas de ejemplo

En el siguiente ejemplo, se recupera información del error de STL_ERROR.

```
select process, errcode, linenum as line,
trim(error) as err
from stl_error;

   process    | errcode | line |                                 err
--------------+---------+------
+-----------------------------------------------------------------
 padbmaster   |    8001 |  194 | Path prefix: s3://redshift-downloads/testnulls/
venue.txt*
 padbmaster   |    8001 |  529 | Listing bucket=redshift-downloads prefix=tests/
category-csv-quotes
 padbmaster   |       2 |  190 | database "template0" is not currently accepting
 connections
 padbmaster   |      32 | 1956 | pq_flush: could not send data to client: Broken pipe
(4 rows)
```

## STL_EXPLAIN

Muestra el plan EXPLAIN de una consulta que se ha enviado para ejecutarse.

STL_EXPLAIN es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_EXPLAIN solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_QUERY_EXPLAIN. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| nodeid | entero | Identificador del nodo del plan, donde se asigna un nodo a uno o más pasos en la ejecución de la consulta. |
| parentid | entero | Identificador del nodo del plan para un nodo principal. Un nodo principal tiene algunos nodos secundarios. Por ejemplo, una combinación de fusión es el nodo principal de los exámenes de las tablas combinadas. |
| plannode | character (400) | El texto del nodo extraído de la salida de EXPLAIN. Los nodos del plan que hacen referencia a la ejecución en nodos de computación llevan el prefijo **XN** en la salida de EXPLAIN. |
| info | character(400) | Información del calificador y del filtro para el nodo del plan. Por ejemplo, las condiciones de combinación y las restricci ones de la cláusula WHERE están incluidas en esta columna. |

## Consultas de ejemplo

Analice la siguiente salida de EXPLAIN para una consulta de combinación de agregación:

```
explain select avg(datediff(day, listtime, saletime)) as avgwait
from sales, listing where sales.listid = listing.listid;
                                QUERY PLAN

------------------------------------------------------------------------------
 XN Aggregate  (cost=6350.30..6350.31 rows=1 width=16)
  ->  XN Hash Join DS_DIST_NONE  (cost=47.08..6340.89 rows=3766 width=16)
        Hash Cond: ("outer".listid = "inner".listid)
        -> XN Seq Scan on listing  (cost=0.00..1924.97 rows=192497 width=12)
        -> XN Hash  (cost=37.66..37.66 rows=3766 width=12)
              -> XN Seq Scan on sales  (cost=0.00..37.66 rows=3766 width=12)
(6 rows)
```

Si ejecuta esta consulta y su ID de consulta es 10, puede utilizar la tabla STL_EXPLAIN para ver el mismo tipo de información que devuelve el comando EXPLAIN:

```
select query,nodeid,parentid,substring(plannode from 1 for 30),
substring(info from 1 for 20) from stl_explain
where query=10 order by 1,2;

query| nodeid |parentid|           substring            |    substring
-----+--------+--------+--------------------------------+-------------------
10   |      1 |      0 |XN Aggregate  (cost=6717.61..6  |
10   |      2 |      1 |  -> XN Merge Join DS_DIST_NO   | Merge Cond:("outer"
10   |      3 |      2 |       -> XN Seq Scan on lis    |
10   |      4 |      2 |       -> XN Seq Scan on sal    |
(4 rows)
```

Analice la siguiente consulta:

```
select event.eventid, sum(pricepaid)
from event, sales
where event.eventid=sales.eventid
group by event.eventid order by 2 desc;

eventid |   sum
--------+----------
    289 | 51846.00
```

```
   7895 | 51049.00
   1602 | 50301.00
    851 | 49956.00
   7315 | 49823.00
...
```

Si el ID de esta consulta es 15, la siguiente consulta a la vista del sistema devolverá los nodos del plan que se completaron. En este caso, el orden de los nodos está invertido para mostrar el orden rea de la ejecución:

```
select query,nodeid,parentid,substring(plannode from 1 for 56)
from stl_explain where query=15 order by 1, 2 desc;

query|nodeid|parentid|                          substring
-----+------+--------+-----------------------------------------------------------
15   |    8 |      7 |                                -> XN Seq Scan on eve
15   |    7 |      5 |                          -> XN Hash(cost=87.98..87.9
15   |    6 |      5 |                          -> XN Seq Scan on sales(cos
15   |    5 |      4 |                    -> XN Hash Join DS_DIST_OUTER(cos
15   |    4 |      3 |              -> XN HashAggregate(cost=862286577.07..
15   |    3 |      2 |        -> XN Sort(cost=1000862287175.47..10008622871
15   |    2 |      1 |  -> XN Network(cost=1000862287175.47..1000862287197.
15   |    1 |      0 |XN Merge(cost=1000862287175.47..1000862287197.46 rows=87
(8 rows)
```

La siguiente consulta recupera los ID de consultas de todos los planes de consulta que tienen una función de ventana:

```
select query, trim(plannode) from stl_explain
where plannode like '%Window%';

query|                                  btrim
-----+-------------------------------------------------------------------------
26   | -> XN Window(cost=1000985348268.57..1000985351256.98 rows=170 width=33)
27   | -> XN Window(cost=1000985348268.57..1000985351256.98 rows=170 width=33)
(2 rows)
```

## STL_FILE_SCAN

Devuelve los archivos que lee Amazon Redshift mientras carga los datos a través del comando COPY.

Consultar esta vista puede ayudar a solucionar los errores de carga de datos. STL_FILE_SCAN puede ser muy útil para identificar problemas en las cargas de datos en paralelo porque, por lo general, las cargas de datos en paralelo cargan varios archivos con un mismo comando COPY.

STL_FILE_SCAN es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_FILE_SCAN solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_LOAD_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| slice | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |
| nombre | character(90) | Ruta y nombre completos del archivo que se cargó. |
| lines | bigint | Cantidad de líneas que se leyeron del archivo. |
| bytes | bigint | Cantidad de bytes que se leyeron del archivo. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| loadtime | bigint | Cantidad de tiempo empleado en la carga del archivo (en microsegundos). |
| curtime | Timestamp | Marca temporal que representa la hora en la que Amazon Redshift comenzó a procesar el archivo. |
| is_partial | entero | Valor que, si es true (1), indica que el archivo de entrada se divide en rangos durante una operación COPY. Si este valor es false (0), el archivo de entrada no se divide. |
| start_offset | bigint | Valor que, si el archivo de entrada se divide durante una operación COPY, indica el valor de desplazamiento de la división (en bytes). Si el archivo no se divide, este valor es 0. |

### Consultas de ejemplo

La siguiente consulta recupera los nombres y los tiempos de carga de todos los archivos que a Amazon Redshift le tomó más de 1 000 000 de microsegundos leer.

```
select trim(name)as name, loadtime from stl_file_scan
where loadtime > 1000000;
```

Esta consulta devuelve el siguiente ejemplo de salida.

```
           name             | loadtime
----------------------------+----------
 listings_pipe.txt          |  9458354
 allusers_pipe.txt          |  2963761
 allevents_pipe.txt         |  1409135
 tickit/listings_pipe.txt   |  7071087
 tickit/allevents_pipe.txt  |  1237364
 tickit/allusers_pipe.txt   |  2535138
 listings_pipe.txt          |  6706370
 allusers_pipe.txt          |  3579461
 allevents_pipe.txt         |  1313195
 tickit/allusers_pipe.txt   |  3236060
 tickit/listings_pipe.txt   |  4980108
```

```
(11 rows)
```

## STL_HASH

Analiza los pasos de ejecución hash de las consultas.

STL_HASH es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_HASH solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| slice | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |
| segment | entero | Número que identifica el segmento de la consulta. |
| paso | entero | Paso de la consulta que se ejecutó. |
| starttime | marca de tiempo | Hora en UTC en que comenzó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | para fracciones de segundo. Por ejemplo: **2009-06-12 11:29:19.131358** . |
| endtime | marca de tiempo | Hora en UTC en la que finalizó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **2009-06-12 11:29:19.131358** . |
| tasknum | entero | Es el número del proceso de tarea de la consulta que se asignó para ejecutar el paso. |
| rows | bigint | Es la cantidad total de filas que se procesaron. |
| bytes | bigint | Tamaño, en bytes, de todas las filas de resultado del paso. |
| slots | entero | Cantidad total de buckets hash. |
| occupied | entero | Cantidad total de slots que tienen registros. |
| maxlength | entero | Tamaño del slot más grande. |
| tbl | entero | ID de la tabla. |
| is_diskbased | character (1) | Si es true (t) (verdadero), significa que la consulta se realizó como una operación basada en el disco. Si es false (f) (falso), implica que la consulta se realizó en la memoria. |
| workmem | bigint | Cantidad total de bytes de memoria funcional asignada a este paso. |
| num_parts | entero | Cantidad total de particiones en que se dividió una tabla hash durante un paso hash. |
| est_rows | bigint | Cantidad estimada de filas a la que se le aplicará una función hash. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| num_blocks_permitted | entero | Esta información es solo para uso interno. |
| resizes | entero | Esta información es solo para uso interno. |
| checksum | bigint | Esta información es solo para uso interno. |
| runtime_filter_size | entero | Tamaño del filtro en tiempo de ejecución, en bytes. |
| max_runtime_filter_size | entero | Tamaño máximo del filtro en tiempo de ejecución, en bytes. |

### Consultas de ejemplo

El siguiente ejemplo devuelve información acerca de la cantidad de particiones que se utilizaron en una función hash para la consulta 720, y se indica que ningún paso se ejecutó en el disco.

```
select slice, rows, bytes, occupied, workmem, num_parts, est_rows,
 num_blocks_permitted, is_diskbased
from stl_hash
where query=720 and segment=5
order by slice;
```

```
 slice | rows | bytes  | occupied | workmem  | num_parts | est_rows |
 num_blocks_permitted | is_diskbased
-------+------+--------+----------+----------+-----------+----------
+----------------------+--------------
     0 |  145 | 585800 |        1 | 88866816 |        16 |        1 |
 52                   f
     1 |    0 |      0 |        0 |        0 |        16 |        1 |
 52                   f
(2 rows)
```

# STL_HASHJOIN

Analiza los pasos de ejecución de combinación de las consultas.

STL_HASHJOIN es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_HASHJOIN solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| slice | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |
| segment | entero | Número que identifica el segmento de la consulta. |
| paso | entero | Paso de la consulta que se ejecutó. |
| starttime | marca de tiempo | Hora en UTC en que comenzó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| endtime | marca de tiempo | Hora en UTC en la que finalizó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| tasknum | entero | Es el número del proceso de tarea de la consulta que se asignó para ejecutar el paso. |
| rows | bigint | Es la cantidad total de filas que se procesaron. |
| tbl | entero | ID de la tabla. |
| num_parts | entero | Cantidad total de particiones en que se dividió una tabla hash durante un paso hash. |
| join_type | entero | Tipo de combinación del paso:<br>• 0. La consulta utilizó una combinación interna.<br>• 1. La consulta utilizó una combinación externa izquierda.<br>• 2. La consulta utilizó una combinación externa completa.<br>• 3. La consulta utilizó una combinación externa derecha.<br>• 4. La consulta utilizó un operador UNION.<br>• 5. La consulta utilizó una condición IN.<br>• 6. Esta información es solo para uso interno.<br>• 7. Esta información es solo para uso interno.<br>• 8. Esta información es solo para uso interno.<br>• 9. Esta información es solo para uso interno.<br>• 10. Esta información es solo para uso interno.<br>• 11. Esta información es solo para uso interno.<br>• 12. Esta información es solo para uso interno. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| hash_looped | character (1) | Esta información es solo para uso interno. |
| switched_parts | character (1) | Indica si los extremos de compilación (externa) y de sonda (interna) se alternaron. |
| used_prefetching | character (1) | Esta información es solo para uso interno. |
| hash_segment | entero | Segmento del paso hash correspondiente. |
| hash_step | entero | Número del paso hash correspondiente. |
| checksum | bigint | Esta información es solo para uso interno. |
| distribución | entero | Esta información es solo para uso interno. |

Consultas de ejemplo

El siguiente ejemplo devuelve la cantidad de particiones utilizadas en una combinación hash para la consulta 720.

```
select query, slice, tbl, num_parts
from stl_hashjoin
where query=720 limit 10;
```

```
 query | slice | tbl | num_parts
-------+-------+-----+-----------
   720 |     0 | 243 |         1
   720 |     1 | 243 |         1
(2 rows)
```

## STL_INSERT

Analiza los pasos de ejecución de inserción de las consultas.

STL_INSERT es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_INSERT solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| slice | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |
| segment | entero | Número que identifica el segmento de la consulta. |
| paso | entero | Paso de la consulta que se ejecutó. |
| starttime | marca de tiempo | Hora en UTC en que comenzó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| endtime | marca de tiempo | Hora en UTC en la que finalizó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | para fracciones de segundo. Por ejemplo: **2009-06-12 11:29:19.131358** . |
| tasknum | entero | Es el número del proceso de tarea de la consulta que se asignó para ejecutar el paso. |
| rows | bigint | Es la cantidad total de filas que se procesaron. |
| tbl | entero | ID de la tabla. |
| inserted_mega_value | character (1) | Esta información es solo para uso interno. Esta información muestra si el paso de insertar proporcionado ha insertado un valor grande. Un valor grande se almacenará en varios bloques. El tamaño del bloque es de 1 MB de forma predeterminada, un valor grande es superior a 1 MB en la configuración predeterminada. |

Consultas de ejemplo

El siguiente ejemplo devuelve los pasos de ejecución de inserción para la consulta más reciente.

```
select slice, segment, step, tasknum, rows, tbl
from stl_insert
where query=pg_last_query_id();
```

```
 slice | segment | step | tasknum | rows  |  tbl
-------+---------+------+---------+-------+--------
     0 |       2 |    2 |      15 | 24958 | 100548
     1 |       2 |    2 |      15 | 25032 | 100548
(2 rows)
```

## STL_LIMIT

Analiza los pasos de ejecución que tienen lugar cuando se utiliza una cláusula LIMIT en una consulta SELECT.

STL_LIMIT es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_LIMIT solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| slice | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |
| segment | entero | Número que identifica el segmento de la consulta. |
| paso | entero | Paso de la consulta que se ejecutó. |
| starttime | marca de tiempo | Hora en UTC en que comenzó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| endtime | marca de tiempo | Hora en UTC en la que finalizó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | para fracciones de segundo. Por ejemplo: **2009-06-12 11:29:19.131358** . |
| tasknum | entero | Es el número del proceso de tarea de la consulta que se asignó para ejecutar el paso. |
| rows | bigint | Es la cantidad total de filas que se procesaron. |
| checksum | bigint | Esta información es solo para uso interno. |

Consultas de ejemplo

Para generar una fila en STL_LIMIT, este ejemplo primero ejecuta la siguiente consulta en la tabla VENUE con la cláusula LIMIT.

```
select * from venue
order by 1
limit 10;
```

```
 venueid |         venuename          |    venuecity    | venuestate | venueseats
---------+----------------------------+-----------------+------------+------------
       1 | Toyota Park                | Bridgeview      | IL         |          0
       2 | Columbus Crew Stadium      | Columbus        | OH         |          0
       3 | RFK Stadium                | Washington      | DC         |          0
       4 | CommunityAmerica Ballpark  | Kansas City     | KS         |          0
       5 | Gillette Stadium           | Foxborough      | MA         |      68756
       6 | New York Giants Stadium    | East Rutherford | NJ         |      80242
       7 | BMO Field                  | Toronto         | ON         |          0
       8 | The Home Depot Center      | Carson          | CA         |          0
       9 | Dick's Sporting Goods Park | Commerce City   | CO         |          0
      10 | Pizza Hut Park             | Frisco          | TX         |          0
(10 rows)
```

Luego, ejecuta la siguiente consulta para encontrar el ID de consulta de la última consulta que se haya ejecutado en la tabla VENUE.

```
select max(query)
```

```
from stl_query;
```

```
  max
--------
 127128
(1 row)
```

Si lo desea, puede ejecutar la siguiente consulta para comprobar que el ID de consulta corresponde a la consulta LIMIT que ejecutó previamente.

```
select query, trim(querytxt)
from stl_query
where query=127128;
```

```
 query  |                  btrim
--------+-----------------------------------------
 127128 | select * from venue order by 1 limit 10;
(1 row)
```

Por último, ejecute la siguiente consulta para obtener información acerca de la consulta LIMIT de la tabla STL_LIMIT.

```
select slice, segment, step, starttime, endtime, tasknum
from stl_limit
where query=127128
order by starttime, endtime;
```

```
  slice | segment | step |         starttime          |          endtime           |
 tasknum
 -------+---------+------+----------------------------+----------------------------
+---------
      1 |       1 |    3 | 2013-09-06 22:56:43.608114 | 2013-09-06 22:56:43.609383 |
   15
      0 |       1 |    3 | 2013-09-06 22:56:43.608708 | 2013-09-06 22:56:43.609521 |
   15
  10000 |       2 |    2 | 2013-09-06 22:56:43.612506 | 2013-09-06 22:56:43.612668 |
    0
(3 rows)
```

# STL_LOAD_COMMITS

Devuelve información para realizar un seguimiento o solucionar problemas en la carga de datos.

En esta vista, se registra el progreso de cada archivo de datos mientras se carga en una tabla de base de datos.

STL_LOAD_COMMITS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_LOAD_COMMITS solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_LOAD_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| slice | entero | Sector cargado para esta entrada. |
| nombre | character (256) | Valor definido por el sistema. |
| filename | character (256) | Nombre del archivo que del que se hace un seguimiento. |
| byte_offset | entero | Esta información es solo para uso interno. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| lines_scanned | entero | Cantidad de líneas examinadas del archivo de carga. Este número puede no coincidir con el número de filas que se cargan efectivamente. Por ejemplo, la carga puede examinar y, al mismo tiempo, tolerar una serie de registros incorrectos, en función de la opción MAXERROR en el comando COPY. |
| errors | entero | Esta información es solo para uso interno. |
| curtime | marca de tiempo | Hora en que se actualizó por última vez esta entrada. |
| status | entero | Esta información es solo para uso interno. |
| file_format | character(16) | Formato del archivo que se va a cargar. Los valores posibles son los siguientes:<br>• Avro<br>• JSON<br>• ORC<br>• Parquet<br>• Texto |
| is_partial | entero | Valor que, si es true (1), indica que el archivo de entrada se divide en rangos durante una operación COPY. Si este valor es false (0), el archivo de entrada no se divide. |
| start_offset | bigint | Valor que, si el archivo de entrada se divide durante una operación COPY, indica el valor de desplazamiento de la división (en bytes). Cada división de archivos se registra como un registro independiente con el valor start_offset correspondiente. Si el archivo no se divide, este valor es 0. |
| copy_job_id | bigint | Identificador del trabajo de copia. Un `0` indica que no hay ningún identificador de trabajo. |

### Consultas de ejemplo

El siguiente ejemplo devuelve los detalles de la última operación COPY.

```
select query, trim(filename) as file, curtime as updated
from stl_load_commits
where query = pg_last_copy_id();

 query |               file               |          updated
-------+----------------------------------+----------------------------
 28554 | s3://dw-tickit/category_pipe.txt | 2013-11-01 17:14:52.648486
(1 row)
```

La siguiente consulta tiene entradas para una carga nueva de las tablas en la base de datos TICKIT:

```
select query, trim(filename), curtime
from stl_load_commits
where filename like '%tickit%' order by query;
```

```
 query |           btrim           |          curtime
-------+---------------------------+----------------------------
 22475 | tickit/allusers_pipe.txt  | 2013-02-08 20:58:23.274186
 22478 | tickit/venue_pipe.txt     | 2013-02-08 20:58:25.070604
 22480 | tickit/category_pipe.txt  | 2013-02-08 20:58:27.333472
 22482 | tickit/date2008_pipe.txt  | 2013-02-08 20:58:28.608305
 22485 | tickit/allevents_pipe.txt | 2013-02-08 20:58:29.99489
 22487 | tickit/listings_pipe.txt  | 2013-02-08 20:58:37.632939
 22593 | tickit/allusers_pipe.txt  | 2013-02-08 21:04:08.400491
 22596 | tickit/venue_pipe.txt     | 2013-02-08 21:04:10.056055
 22598 | tickit/category_pipe.txt  | 2013-02-08 21:04:11.465049
 22600 | tickit/date2008_pipe.txt  | 2013-02-08 21:04:12.461502
 22603 | tickit/allevents_pipe.txt | 2013-02-08 21:04:14.785124
 22605 | tickit/listings_pipe.txt  | 2013-02-08 21:04:20.170594

(12 rows)
```

El hecho de que se escriba un registro en el archivo de registros para esta vista de sistema no implica que la carga se haya confirmado correctamente como parte de la transacción que contiene. Para comprobar las confirmaciones de carga, consulte la vista STL_UTILITYTEXT y busque el registro COMMIT que corresponda a esa transacción COPY. Por ejemplo, esta consulta combina STL_LOAD_COMMITS y STL_QUERY a partir de una subconsulta de STL_UTILITYTEXT:

```
select l.query,rtrim(l.filename),q.xid
from stl_load_commits l, stl_query q
where l.query=q.query
and exists
(select xid from stl_utilitytext where xid=q.xid and rtrim("text")='COMMIT');

 query |           rtrim           |  xid
-------+---------------------------+-------
 22600 | tickit/date2008_pipe.txt  | 68311
 22480 | tickit/category_pipe.txt  | 68066
  7508 | allusers_pipe.txt         | 23365
  7552 | category_pipe.txt         | 23415
  7576 | allevents_pipe.txt        | 23429
  7516 | venue_pipe.txt            | 23390
  7604 | listings_pipe.txt         | 23445
 22596 | tickit/venue_pipe.txt     | 68309
 22605 | tickit/listings_pipe.txt  | 68316
 22593 | tickit/allusers_pipe.txt  | 68305
 22485 | tickit/allevents_pipe.txt | 68071
  7561 | allevents_pipe.txt        | 23429
  7541 | category_pipe.txt         | 23415
  7558 | date2008_pipe.txt         | 23428
 22478 | tickit/venue_pipe.txt     | 68065
   526 | date2008_pipe.txt         |  2572
  7466 | allusers_pipe.txt         | 23365
 22482 | tickit/date2008_pipe.txt  | 68067
 22598 | tickit/category_pipe.txt  | 68310
 22603 | tickit/allevents_pipe.txt | 68315
 22475 | tickit/allusers_pipe.txt  | 68061
   547 | date2008_pipe.txt         |  2572
 22487 | tickit/listings_pipe.txt  | 68072
  7531 | venue_pipe.txt            | 23390
  7583 | listings_pipe.txt         | 23445
(25 rows)
```

En los ejemplos siguientes, se destacan los valores de columna is_partial y start_offset.

```
-- Single large file copy without scan range
SELECT count(*) FROM stl_load_commits WHERE query = pg_last_copy_id();
1

-- Single large uncompressed, delimited file copy with scan range
SELECT count(*) FROM stl_load_commits WHERE query = pg_last_copy_id();
```

```
16

-- Scan range offset logging in the file at 64MB boundary.
SELECT start_offset FROM stl_load_commits
WHERE query = pg_last_copy_id() ORDER BY start_offset;
0
67108864
134217728
201326592
268435456
335544320
402653184
469762048
536870912
603979776
671088640
738197504
805306368
872415232
939524096
1006632960
```

## STL_LOAD_ERRORS

Muestra los registros de todos los errores de carga de Amazon Redshift.

STL_LOAD_ERRORS presenta un historial de todos los errores de carga de Amazon Redshift. Para obtener una lista completa de posibles errores y explicaciones de cargas, consulte Referencia de error de carga.

Consulte STL_LOADERROR_DETAIL para obtener detalles adicionales, como la fila y la columna de datos exactas en las que se produjo un error de análisis, después de consultar STL_LOAD_ERRORS para encontrar información general sobre el error.

STL_LOAD_ERRORS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_LOAD_ERRORS solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado

simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_LOAD_ERROR_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| slice | entero | Sector en el que se produjo el error. |
| tbl | entero | ID de la tabla. |
| starttime | marca de tiempo | La hora de inicio de la carga en UTC. |
| sesión | entero | ID de la sesión que realiza la carga. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| filename | character (256) | Ruta completa hacia el archivo de entrada para la carga. |
| line_number | bigint | Número de línea en el archivo de carga que tiene el error. Para el comando COPY de JSON, el número de línea de la última línea del objeto JSON que tiene el error. |
| colname | character (127) | Campo que tiene el error. |
| type | character (10) | Tipo de datos del campo. |
| col_length | character (10) | Longitud de la columna, si corresponde. Este campo se rellena cuando el tipo de datos tiene una longitud limitada. Por ejemplo, para una columna con un tipo de datos "character (3)", esta columna tendrá el valor "3". |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| position | entero | Posición del error en el campo. |
| raw_line | character (1024) | Datos de carga sin formato que tienen el error. Caracteres multibyte en la carga de datos que se reemplazan con un punto. |
| raw_field_value | char (1024) | El valor antes del análisis del campo "colname" que induce el error de análisis. |
| err_code | entero | Código de error. |
| err_reason | character (100) | Explicación del error. |
| is_partial | entero | Valor que, si es true (1), indica que el archivo de entrada se divide en rangos durante una operación COPY. Si este valor es false (0), el archivo de entrada no se divide. |
| start_offset | bigint | Valor que, si el archivo de entrada se divide durante una operación COPY, indica el valor de desplazamiento de la división (en bytes). Si se desconoce el número de línea del archivo, el número de línea es -1. Si el archivo no se divide, este valor es 0. |
| copy_job_id | bigint | Identificador del trabajo de copia. Un `0` indica que no hay ningún identificador de trabajo. |

Consultas de ejemplo

La siguiente consulta combina STL_LOAD_ERRORS y STL_LOADERROR_DETAIL para ver los errores de detalles que se produjeron durante la carga más reciente.

```
select d.query, substring(d.filename,14,20),
d.line_number as line,
substring(d.value,1,16) as value,
substring(le.err_reason,1,48) as err_reason
from stl_loaderror_detail d, stl_load_errors le
```

```
where d.query = le.query
and d.query = pg_last_copy_id();

 query |    substring      | line |  value   |             err_reason
-------+-------------------+------+----------+----------------------------
    558| allusers_pipe.txt |  251 | 251      | String contains invalid or
                                               unsupported UTF8 code
    558| allusers_pipe.txt |  251 | ZRU29FGR | String contains invalid or
                                               unsupported UTF8 code
    558| allusers_pipe.txt |  251 | Kaitlin  | String contains invalid or
                                               unsupported UTF8 code
    558| allusers_pipe.txt |  251 | Walter   | String contains invalid or
                                               unsupported UTF8 code
```

En el siguiente ejemplo, se usa STL_LOAD_ERRORS con STV_TBL_PERM para crear una vista nueva y, luego, se usa esa vista para determinar qué errores se produjeron mientras se cargaban los datos en la tabla EVENT:

```
create view loadview as
(select distinct tbl, trim(name) as table_name, query, starttime,
trim(filename) as input, line_number, colname, err_code,
trim(err_reason) as reason
from stl_load_errors sl, stv_tbl_perm sp
where sl.tbl = sp.id);
```

Luego, la consulta siguiente devuelve efectivamente el último error que se produjo mientras se cargaba la tabla EVENT:

```
select table_name, query, line_number, colname, starttime,
trim(reason) as error
from loadview
where table_name ='event'
order by line_number limit 1;
```

La consulta devuelve el último error de carga que se produjo para la tabla EVENT. Si no se produjeron errores de carga, la consulta devuelve cero filas. En este ejemplo, la consulta devuelve un error único:

```
 table_name | query | line_number | colname | error | starttime
------+-----+----+----+-------------------------------------------------------
+----------------------
```

```
event | 309 |  0 |  5 | Error in Timestamp value or format [%Y-%m-%d %H:%M:%S] |
 2014-04-22 15:12:44

(1 row)
```

En los casos en que el comando COPY divide automáticamente datos de archivos grandes, sin comprimir y delimitados por texto para facilitar el paralelismo, las columnas line_number, is_partial y start_offset muestran información relativa a las divisiones. (El número de línea puede ser desconocido en los casos en que el número de línea del archivo original no esté disponible).

```
--scan ranges information
SELECT line_number, POSITION, btrim(raw_line), btrim(raw_field_value),
btrim(err_reason), is_partial, start_offset FROM stl_load_errors
WHERE query = pg_last_copy_id();

--result
-1,51,"1008771|13463413|463414|2|28.00|38520.72|0.06|0.07|NO|1998-08-30|1998-09-25|
1998-09-04|TAKE BACK RETURN|RAIL|ans cajole sly","NO","Char length exceeds DDL
 length",1,67108864
```

## STL_LOADERROR_DETAIL

Muestra un registro de los errores de análisis de datos que ocurrieron mientras se usaba el comando COPY para cargar las tablas. Para conservar espacio en el disco, se registran 20 errores como máximo por sector de nodo para cada operación de carga.

Un error de análisis ocurre cuando Amazon Redshift no puede analizar un campo en una fila de datos mientras lo carga en una tabla. Por ejemplo, si la columna de una tabla está esperando un tipo de datos entero y el archivo de datos tiene una cadena de letras en ese campo, se produce un error de análisis.

Consulte STL_LOADERROR_DETAIL para obtener información adicional, como la fila y la columna de datos exactas en la que ocurrió un error de análisis, después de consultar STL_LOAD_ERRORS para averiguar información general sobre el error.

La vista STL_LOADERROR_DETAIL tiene todas las columnas de datos, incluida la columna previa a aquella en la que ocurrió el error de análisis. Utilice el campo VALUE para ver el valor de los datos que realmente se analizaron en esta columna, incluidas las columnas que se analizaron correctamente hasta el error.

Todos los usuarios pueden acceder a esta vista. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_LOADERROR_DETAIL solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_LOAD_ERROR_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| slice | entero | Sector en el que se produjo el error. |
| sesión | entero | ID de la sesión que realiza la carga. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| filename | character (256) | Ruta completa hacia el archivo de entrada para la carga. |
| line_number | bigint | Número de línea en el archivo de carga que tiene el error. |
| field | entero | Campo que tiene el error. |
| colname | character(1024) | Nombre de la columna. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| valor | character(1024) | Valor del dato analizado del campo. (puede estar truncado). Caracteres multibyte en la carga de datos que se reemplaza n con un punto. |
| is_null | entero | Indica si el valor analizado es nulo o no. |
| type | character (10) | Tipo de datos del campo. |
| col_length | character (10) | Longitud de la columna, si corresponde. Este campo se rellena cuando el tipo de datos tiene una longitud limitada. Por ejemplo, para una columna con un tipo de datos "character (3)", esta columna tendrá el valor "3". |

Consulta de ejemplo

La siguiente consulta combina STL_LOAD_ERRORS y STL_LOADERROR_DETAIL para ver los detalles de un error de análisis que ocurrió mientras se cargaba la tabla EVENT, que tiene un ID de tabla de 100133:

```
select d.query, d.line_number, d.value,
le.raw_line, le.err_reason
from stl_loaderror_detail d, stl_load_errors le
where
d.query = le.query
and tbl = 100133;
```

En el siguiente ejemplo de salida, se muestran las columnas que se cargaron correctamente, incluida la columna con el error. En este ejemplo, se cargaron dos columnas correctamente antes de que ocurriera el error de análisis en la tercera columna, donde se analizó una cadena de caracteres incorrectamente para un campo que esperaba un valor entero. Como el campo esperaba un valor entero, analizó la cadena "aaa", que es un dato sin inicializar, como un valor nulo y generó un error de análisis. La salida muestra el valor sin procesar, el valor analizado y el motivo del error:

```
query  | line_number | value | raw_line | err_reason
-------+-------------+-------+----------+----------------
4      |      3      |  1201 |  1201    | Invalid digit
```

```
4      |      3      |   126 |   126    | Invalid digit
4      |      3      |       |   aaa    | Invalid digit
(3 rows)
```

Cuando una consulta combina STL_LOAD_ERRORS y STL_LOADERROR_DETAIL, muestra un motivo de error por cada columna en la fila de datos, que simplemente significa que ocurrió un error en esa fila. La última fila en los resultados es la columna real en la que ocurrió el error.

## STL_MERGE

Analiza los pasos de ejecución de fusión de las consultas. Estos pasos tienen lugar cuando se combinan resultados de operaciones paralelas (como ordenaciones y combinaciones) para un posterior procesamiento.

STL_MERGE es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_MERGE solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| slice | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |
| segment | entero | Número que identifica el segmento de la consulta. |
| paso | entero | Paso de la consulta que se ejecutó. |
| starttime | marca de tiempo | Hora en UTC en que comenzó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **2009-06-12 11:29:19.131358** . |
| endtime | marca de tiempo | Hora en UTC en la que finalizó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **2009-06-12 11:29:19.131358** . |
| tasknum | entero | Es el número del proceso de tarea de la consulta que se asignó para ejecutar el paso. |
| rows | bigint | Es la cantidad total de filas que se procesaron. |

Consultas de ejemplo

El siguiente ejemplo devuelve los resultados de 10 ejecuciones de fusión.

```
select query, step, starttime, endtime, tasknum, rows
from stl_merge
limit 10;
```

```
 query | step |       starttime     |        endtime      | tasknum | rows
-------+------+---------------------+---------------------+---------+------
     9 |    0 | 2013-08-12 20:08:14 | 2013-08-12 20:08:14 |       0 |    0
    12 |    0 | 2013-08-12 20:09:10 | 2013-08-12 20:09:10 |       0 |    0
    15 |    0 | 2013-08-12 20:10:24 | 2013-08-12 20:10:24 |       0 |    0
    20 |    0 | 2013-08-12 20:11:27 | 2013-08-12 20:11:27 |       0 |    0
```

```
    26 |    0 | 2013-08-12 20:12:28 | 2013-08-12 20:12:28 |       0 |    0
    32 |    0 | 2013-08-12 20:14:33 | 2013-08-12 20:14:33 |       0 |    0
    38 |    0 | 2013-08-12 20:16:43 | 2013-08-12 20:16:43 |       0 |    0
    44 |    0 | 2013-08-12 20:17:05 | 2013-08-12 20:17:05 |       0 |    0
    50 |    0 | 2013-08-12 20:18:48 | 2013-08-12 20:18:48 |       0 |    0
    56 |    0 | 2013-08-12 20:20:48 | 2013-08-12 20:20:48 |       0 |    0
(10 rows)
```

## STL_MERGEJOIN

Analiza los pasos de ejecución de combinaciones de fusión de las consultas.

STL_MERGEJOIN es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_MERGEJOIN solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| slice | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| segment | entero | Número que identifica el segmento de la consulta. |
| paso | entero | Paso de la consulta que se ejecutó. |
| starttime | marca de tiempo | Hora en UTC en que comenzó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **2009-06-12 11:29:19.131358** . |
| endtime | marca de tiempo | Hora en UTC en la que finalizó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **2009-06-12 11:29:19.131358** . |
| tasknum | entero | Es el número del proceso de tarea de la consulta que se asignó para ejecutar el paso. |
| rows | bigint | Es la cantidad total de filas que se procesaron. |
| tbl | entero | ID de la tabla. Este es el ID para la tabla interna que se utilizó en la combinación de fusión. |
| checksum | bigint | Esta información es solo para uso interno. |

Consultas de ejemplo

El siguiente ejemplo devuelve los resultados de combinaciones de fusión para la consulta más reciente.

```
select sum(s.qtysold), e.eventname
from event e, listing l, sales s
where e.eventid=l.eventid
and l.listid= s.listid
group by e.eventname;

select * from stl_mergejoin where query=pg_last_query_id();
```

```
 userid | query | slice | segment | step |         starttime   |          endtime    |
 tasknum | rows | tbl
--------+-------+-------+---------+------+---------------------+---------------------
+---------+------+-----
    100 | 27399 |     3 |       4 |    4 | 2013-10-02 16:30:41 | 2013-10-02 16:30:41 |
     19 |43428 | 240
    100 | 27399 |     0 |       4 |    4 | 2013-10-02 16:30:41 | 2013-10-02 16:30:41 |
     19 |43159 | 240
    100 | 27399 |     2 |       4 |    4 | 2013-10-02 16:30:41 | 2013-10-02 16:30:41 |
     19 |42778 | 240
    100 | 27399 |     1 |       4 |    4 | 2013-10-02 16:30:41 | 2013-10-02 16:30:41 |
     19 |43091 | 240
```

## STL_MV_STATE

La vista STL_MV_STATE contiene una fila para cada transición de estado de una vista materializada.

Para obtener más información acerca de las vistas materializadas, consulte Vistas materializadas en Amazon Redshift.

STL_MV_STATE es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_MV_STATE. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | bigint | El ID del usuario que creó el evento. |
| starttime | marca de tiempo | La hora de inicio del evento. |
| xid | bigint | El ID de transacción del evento. |
| event_desc | char(500) | El evento que solicitó el cambio de estado. Algunos valores de ejemplo son los siguientes: |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | • Se ha cambiado el tipo de columna<br>• Se ha eliminado la columna<br>• Se ha cambiado el nombre de la columna<br>• Se ha cambiado el nombre del esquema<br>• Conversión de tabla pequeña<br>• TRUNCATE<br>• Vacuum<br><br>Tenga en cuenta que existen otros valores posibles para esta columna. |
| db_name | char(128) | Base de datos que contiene la vista materializada. |
| base_table_schema | char(128) | Esquema de la tabla base. |
| base_table_name | char(128) | Nombre de la tabla base. |
| mv_schema | char(128) | Esquema de la vista materializada. |
| mv_name | char(128) | El nombre de la vista materializada. |
| estado | character (32) | El estado cambiado de la vista materializada es el siguiente:<br>• Volver a calcular<br>• No se puede actualizar |

En la tabla siguiente se muestran ejemplos de combinaciones de `event_desc` y `state`.

```
         event_desc     |     state
------------------------+--------------
```

```
 TRUNCATE                | Recompute
 TRUNCATE                | Recompute
 Small table conversion  | Recompute
 Vacuum                  | Recompute
 Column was renamed      | Unrefreshable
 Column was dropped      | Unrefreshable
 Table was renamed       | Unrefreshable
 Column type was changed | Unrefreshable
 Schema name was changed | Unrefreshable
```

### Consulta de ejemplo

Para ver el registro de transiciones de estado de vistas materializadas, ejecute la siguiente consulta.

```
select * from stl_mv_state;
```

Esta consulta devuelve el siguiente ejemplo de salida:

```
 userid |         starttime          | xid  |            event_desc       | db_name |
 base_table_schema   |   base_table_name    |      mv_schema       | mv_name       |
  state
--------+----------------------------+------+-----------------------------+---------
+---------------------+---------------------+---------------------+--------------
+---------------
    138 | 2020-02-14 02:21:25.578885 | 5180 | TRUNCATE                    | dev     |
 public              | mv_base_table        | public              | mv_test       |
 Recompute
    138 | 2020-02-14 02:21:56.846774 | 5275 | Column was dropped          | dev     |
                     | mv_base_table        | public              | mv_test       |
 Unrefreshable
    100 | 2020-02-13 22:09:53.041228 | 1794 | Column was renamed          | dev     |
                     | mv_base_table        | public              | mv_test       |
 Unrefreshable
      1 | 2020-02-13 22:10:23.630914 | 1893 | ALTER TABLE ALTER SORTKEY   | dev     |
 public              | mv_base_table_sorted | public              | mv_test       |
 Recompute
      1 | 2020-02-17 22:57:22.497989 | 8455 | ALTER TABLE ALTER DISTSTYLE | dev     |
 public              | mv_base_table        | public              | mv_test       |
 Recompute
    173 | 2020-02-17 22:57:23.591434 | 8504 | Table was renamed           | dev     |
                     | mv_base_table        | public              | mv_test       |
 Unrefreshable
```

```
   173 | 2020-02-17 22:57:27.229423 | 8592 | Column type was changed     | dev     |
                    | mv_base_table        | public               | mv_test       |
Unrefreshable
   197 | 2020-02-17 22:59:06.212569 | 9668 | TRUNCATE                    | dev     |
schemaf796e415850f4f | mv_base_table        | schemaf796e415850f4f | mv_test       |
Recompute
   138 | 2020-02-14 02:21:55.705655 | 5226 | Column was renamed          | dev     |
                    | mv_base_table        | public               | mv_test       |
Unrefreshable
     1 | 2020-02-14 02:22:26.292434 | 5325 | ALTER TABLE ALTER SORTKEY   | dev     |
public               | mv_base_table_sorted | public               | mv_test       |
Recompute
```

## STL_NESTLOOP

Analiza los pasos de ejecución de combinaciones de bucle anidado de las consultas.

STL_NESTLOOP es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_NESTLOOP solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| slice | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |
| segment | entero | Número que identifica el segmento de la consulta. |
| paso | entero | Paso de la consulta que se ejecutó. |
| starttime | marca de tiempo | Hora en UTC en que comenzó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **2009-06-12 11:29:19.131358** . |
| endtime | marca de tiempo | Hora en UTC en la que finalizó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **2009-06-12 11:29:19.131358** . |
| tasknum | entero | Es el número del proceso de tarea de la consulta que se asignó para ejecutar el paso. |
| rows | bigint | Es la cantidad total de filas que se procesaron. |
| tbl | entero | ID de la tabla. |
| checksum | bigint | Esta información es solo para uso interno. |

Consultas de ejemplo

Como la siguiente consulta no contempla la combinación con la tabla CATEGORY, produce un producto cartesiano parcial, que no es recomendable. Aquí se muestra para ilustrar un bucle anidado.

```
select count(event.eventname), event.eventname, category.catname, date.caldate
```

```
from event, category, date
where event.dateid = date.dateid
group by event.eventname, category.catname, date.caldate;
```

La siguiente consulta muestra los resultados de la consulta anterior en la vista STL_NESTLOOP.

```
select query, slice, segment as seg, step,
datediff(msec, starttime, endtime) as duration, tasknum, rows, tbl
from stl_nestloop
where query = pg_last_query_id();
```

```
 query | slice | seg | step | duration | tasknum | rows  | tbl
-------+-------+-----+------+----------+---------+-------+-----
  6028 |     0 |   4 |    5 |       41 |      22 | 24277 | 240
  6028 |     1 |   4 |    5 |       26 |      23 | 24189 | 240
  6028 |     3 |   4 |    5 |       25 |      23 | 24376 | 240
  6028 |     2 |   4 |    5 |       54 |      22 | 23936 | 240
```

## STL_PARSE

Analiza los pasos de la consulta que analizan las cadenas en valores binarios para la carga.

STL_PARSE es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_PARSE solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| slice | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |
| segment | entero | Número que identifica el segmento de la consulta. |
| paso | entero | Paso de la consulta que se ejecutó. |
| starttime | marca de tiempo | Hora en UTC en que comenzó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| endtime | marca de tiempo | Hora en UTC en la que finalizó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| tasknum | entero | Es el número del proceso de tarea de la consulta que se asignó para ejecutar el paso. |
| rows | bigint | Es la cantidad total de filas que se procesaron. |

Consultas de ejemplo

El siguiente ejemplo devuelve todos los resultados de pasos de la consulta para el sector 1 y el segmento 0 donde se analizaron las cadenas en valores binarios.

```
select query, step, starttime, endtime, tasknum, rows
from stl_parse
```

```
where slice=1 and segment=0;
```

```
 query | step |     starttime       |        endtime      | tasknum |  rows
-------+------+---------------------+---------------------+---------+--------
   669 |    1 | 2013-08-12 22:35:13 | 2013-08-12 22:35:17 |      32 | 192497
   696 |    1 | 2013-08-12 22:35:49 | 2013-08-12 22:35:49 |      32 |      0
   525 |    1 | 2013-08-12 22:32:03 | 2013-08-12 22:32:03 |      13 |  49990
   585 |    1 | 2013-08-12 22:33:18 | 2013-08-12 22:33:19 |      13 |    202
   621 |    1 | 2013-08-12 22:34:03 | 2013-08-12 22:34:03 |      27 |    365
   651 |    1 | 2013-08-12 22:34:47 | 2013-08-12 22:34:53 |      35 | 192497
   590 |    1 | 2013-08-12 22:33:28 | 2013-08-12 22:33:28 |      19 |      0
   599 |    1 | 2013-08-12 22:33:39 | 2013-08-12 22:33:39 |      31 |     11
   675 |    1 | 2013-08-12 22:35:26 | 2013-08-12 22:35:27 |      38 |   3766
   567 |    1 | 2013-08-12 22:32:47 | 2013-08-12 22:32:48 |      23 |  49990
   630 |    1 | 2013-08-12 22:34:17 | 2013-08-12 22:34:17 |      36 |      0
   572 |    1 | 2013-08-12 22:33:04 | 2013-08-12 22:33:04 |      29 |      0
   645 |    1 | 2013-08-12 22:34:37 | 2013-08-12 22:34:38 |      29 |   8798
   604 |    1 | 2013-08-12 22:33:47 | 2013-08-12 22:33:47 |      37 |      0
(14 rows)
```

## STL_PLAN_INFO

Utilice la vista STL_PLAN_INFO para analizar la salida del comando EXPLAIN para una consulta en términos de conjunto de filas. Esta es una manera alternativa de analizar los planes de consulta.

STL_PLAN_INFO es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_PLAN_INFO solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| nodeid | entero | Identificador del nodo del plan, donde se asigna un nodo a uno o más pasos en la ejecución de la consulta. |
| segment | entero | Número que identifica el segmento de la consulta. |
| paso | entero | Número que identifica el paso de la consulta. |
| locus | entero | Ubicación en la que se ejecuta el paso. 0 si es en un nodo de informática y 1 si es en el nodo principal. |
| plannode | entero | Valor enumerado del nodo del plan. Para ver enums de la columna plannode, consulte la siguiente tabla. (La columna PLANNODE en STL_EXPLAIN tiene el texto del nodo del plan). |
| startupcost | double precision | El costo relativo estimado de devolver la primera fila de este paso. |
| totalcost | double precision | El costo relativo estimado de ejecutar este paso. |
| rows | bigint | La cantidad estimada de filas que producirá el paso. |
| bytes | bigint | La cantidad estimada de bytes que producirá el paso. |

Consultas de ejemplo

En los siguientes ejemplos, se comparan los planes de consulta para una consulta simple SELECT, que devuelve el uso del comando EXPLAIN y la consulta de la vista STL_PLAN_INFO.

```
explain select * from category;
```

```
QUERY PLAN
------------------------------------------------------------
XN Seq Scan on category (cost=0.00..0.11 rows=11 width=49)
(1 row)

select * from category;
catid | catgroup | catname | catdesc
-------+----------+-----------+--------------------------------------------
1 | Sports | MLB | Major League Baseball
3 | Sports | NFL | National Football League
5 | Sports | MLS | Major League Soccer
...

select * from stl_plan_info where query=256;

query | nodeid | segment | step | locus | plannode | startupcost | totalcost
| rows | bytes
-------+--------+---------+------+-------+----------+-------------+-----------+------
+-------
256 | 1 | 0 | 1 | 0 | 104 | 0 | 0.11 | 11 | 539
256 | 1 | 0 | 0 | 0 | 104 | 0 | 0.11 | 11 | 539
(2 rows)
```

En este ejemplo, PLANNODE 104 hace referencia al examen secuencial de la tabla CATEGORY.

```
select distinct eventname from event order by 1;

eventname
----------------------------------------------------------------------------
.38 Special
3 Doors Down
70s Soul Jam
A Bronx Tale
...

explain select distinct eventname from event order by 1;

QUERY PLAN
------------------------------------------------------------------------------------------
XN Merge (cost=1000000000136.38..1000000000137.82 rows=576 width=17)
Merge Key: eventname
-> XN Network (cost=1000000000136.38..1000000000137.82 rows=576
width=17)
```

```
Send to leader
-> XN Sort (cost=1000000000136.38..1000000000137.82 rows=576
width=17)
Sort Key: eventname
-> XN Unique (cost=0.00..109.98 rows=576 width=17)
-> XN Seq Scan on event (cost=0.00..87.98 rows=8798
width=17)
(8 rows)

select * from stl_plan_info where query=240 order by nodeid desc;

query | nodeid | segment | step | locus | plannode | startupcost |
totalcost | rows | bytes
-------+--------+---------+------+-------+----------+-----------------
+------------------+------+--------
240 | 5 | 0 | 0 | 0 | 104 | 0                | 87.98   | 8798 | 149566
240 | 5 | 0 | 1 | 0 | 104 | 0                | 87.98   | 8798 | 149566
240 | 4 | 0 | 2 | 0 | 117 | 0                | 109.975 | 576  | 9792
240 | 4 | 0 | 3 | 0 | 117 | 0                | 109.975 | 576  | 9792
240 | 4 | 1 | 0 | 0 | 117 | 0                | 109.975 | 576  | 9792
240 | 4 | 1 | 1 | 0 | 117 | 0                | 109.975 | 576  | 9792
240 | 3 | 1 | 2 | 0 | 114 | 1000000000136.38 | 1000000000137.82 | 576 | 9792
240 | 3 | 2 | 0 | 0 | 114 | 1000000000136.38 | 1000000000137.82 | 576 | 9792
240 | 2 | 2 | 1 | 0 | 123 | 1000000000136.38 | 1000000000137.82 | 576 | 9792
240 | 1 | 3 | 0 | 0 | 122 | 1000000000136.38 | 1000000000137.82 | 576 | 9792
(10 rows)
```

## STL_PROJECT

Presenta las filas de los pasos de la consulta que se utilizaron para evaluar expresiones.

STL_PROJECT es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_PROJECT solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión

de SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| slice | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |
| segment | entero | Número que identifica el segmento de la consulta. |
| paso | entero | Paso de la consulta que se ejecutó. |
| starttime | marca de tiempo | Hora en UTC en que comenzó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| endtime | marca de tiempo | Hora en UTC en la que finalizó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| tasknum | entero | Es el número del proceso de tarea de la consulta que se asignó para ejecutar el paso. |
| rows | bigint | Es la cantidad total de filas que se procesaron. |
| checksum | bigint | Esta información es solo para uso interno. |

### Consultas de ejemplo

El siguiente ejemplo devuelve todas las filas de los pasos de la consulta que se utilizaron para evaluar las expresiones del sector 0 y del segmento 1.

```
select query, step, starttime, endtime, tasknum, rows
from stl_project
where slice=0 and segment=1;
```

```
 query  | step |         starttime   |          endtime    | tasknum | rows
--------+------+---------------------+---------------------+---------+------
  86399 |    2 | 2013-08-29 22:01:21 | 2013-08-29 22:01:21 |      25 |   -1
  86399 |    3 | 2013-08-29 22:01:21 | 2013-08-29 22:01:21 |      25 |   -1
    719 |    1 | 2013-08-12 22:38:33 | 2013-08-12 22:38:33 |       7 |   -1
  86383 |    1 | 2013-08-29 21:58:35 | 2013-08-29 21:58:35 |       7 |   -1
    714 |    1 | 2013-08-12 22:38:17 | 2013-08-12 22:38:17 |       2 |   -1
  86375 |    1 | 2013-08-29 21:57:59 | 2013-08-29 21:57:59 |       2 |   -1
  86397 |    2 | 2013-08-29 22:01:20 | 2013-08-29 22:01:20 |      19 |   -1
    627 |    1 | 2013-08-12 22:34:13 | 2013-08-12 22:34:13 |      34 |   -1
  86326 |    2 | 2013-08-29 21:45:28 | 2013-08-29 21:45:28 |      34 |   -1
  86326 |    3 | 2013-08-29 21:45:28 | 2013-08-29 21:45:28 |      34 |   -1
  86325 |    2 | 2013-08-29 21:45:27 | 2013-08-29 21:45:27 |      28 |   -1
  86371 |    1 | 2013-08-29 21:57:42 | 2013-08-29 21:57:42 |       4 |   -1
 111100 |    2 | 2013-09-03 19:04:45 | 2013-09-03 19:04:45 |      12 |   -1
    704 |    2 | 2013-08-12 22:36:34 | 2013-08-12 22:36:34 |      37 |   -1
    649 |    2 | 2013-08-12 22:34:47 | 2013-08-12 22:34:47 |      38 |   -1
    649 |    3 | 2013-08-12 22:34:47 | 2013-08-12 22:34:47 |      38 |   -1
    632 |    2 | 2013-08-12 22:34:22 | 2013-08-12 22:34:22 |      13 |   -1
    705 |    2 | 2013-08-12 22:36:48 | 2013-08-12 22:36:49 |      13 |   -1
    705 |    3 | 2013-08-12 22:36:48 | 2013-08-12 22:36:49 |      13 |   -1
      3 |    1 | 2013-08-12 20:07:40 | 2013-08-12 20:07:40 |       3 |   -1
  86373 |    1 | 2013-08-29 21:57:58 | 2013-08-29 21:57:58 |       3 |   -1
 107976 |    1 | 2013-09-03 04:05:12 | 2013-09-03 04:05:12 |       3 |   -1
  86381 |    1 | 2013-08-29 21:58:35 | 2013-08-29 21:58:35 |       8 |   -1
  86396 |    1 | 2013-08-29 22:01:20 | 2013-08-29 22:01:20 |      15 |   -1
    711 |    1 | 2013-08-12 22:37:10 | 2013-08-12 22:37:10 |      20 |   -1
  86324 |    1 | 2013-08-29 21:45:27 | 2013-08-29 21:45:27 |      24 |   -1
(26 rows)
```

## STL_QUERY

Devuelve información acerca de la ejecución de la consulta de una base de datos.

> **Note**
>
> Las vistas STL_QUERY y STL_QUERYTEXT solo presentan información relacionada con consultas, no sobre otras utilidades y comandos de DDL. Para obtener una lista con información sobre todas las instrucciones ejecutadas por Amazon Redshift, también puede consultar las vistas STL_DDLTEXT y STL_UTILITYTEXT. Para obtener una lista completa de todas las instrucciones ejecutadas por Amazon Redshift, puede consultar la vista SVL_STATEMENTTEXT.

STL_QUERY es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_QUERY_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

> **Note**
>
> Para verificar si una transacción que contiene la consulta ejecutada se ha confirmado correctamente, debe realizar una operación de unión entre las tablas del sistema y la tabla `sys_transaction_history`. Por ejemplo:
>
> ```
> SELECT
>     stlq.xid AS transaction_id,
>     stlq.query AS query_id,
>     TRIM(stlq.querytxt) AS query_text,
>     th.status AS transaction_status
> FROM
>     stl_query stlq
> LEFT JOIN
>     sys_transaction_history th ON stlq.xid = th.transaction_id;
> ```

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| etiqueta | character(320) | Ya sea el nombre del archivo utilizado para ejecutar la consulta o una etiqueta definida con un comando SET QUERY_GROUP. Si la consulta no se basa en archivos o si no se establece el parámetro QUERY_GROUP, el valor del campo es `default`. |
| xid | bigint | ID de transacción. |
| pid | entero | ID de proceso. Por lo general, todas las consultas en una sesión se ejecutan en el mismo proceso, por lo que este valor suele permanecer constante si ejecuta una serie de consultas en la misma sesión. Después de determinados eventos internos, Amazon Redshift podría reiniciar una sesión activa y asignar un PID nuevo. Para obtener más información, consulte STL_RESTARTED_SESSIONS. |
| database | character (32) | El nombre de la base de datos al que estaba conectado el usuario cuando se emitió la consulta. |
| querytxt | character (4000) | Texto real de la consulta para la consulta. |
| starttime | marca de tiempo | Hora en UTC en que comenzó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| endtime | marca de tiempo | Hora en UTC en la que finalizó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | para fracciones de segundo. Por ejemplo: **2009-06-12 11:29:19.131358** . |
| aborted | entero | Si el sistema detuvo la consulta o el usuario la canceló, esta columna tendrá el valor **1**. Si la consulta se ejecutó hasta terminarse (e incluso se devolvieron resultados al cliente), esta columna tendrá el valor **0**. Si un cliente se desconecta antes de recibir los resultados, la consulta se marcará como cancelada (**1**), aunque se haya completad o correctamente en el backend. |
| insert_pristine | entero | Indica si las consultas de escritura pueden o pudieron ejecutarse mientras que la consulta actual se ejecuta o ejecutaba: 1 = no se permiten consultas de escritura; 0 = se permiten las consultas de escritura. Esta columna está diseñada para ser utilizada en la depuración. |
| concurrency_scaling_status | entero | Indica si la consulta se ejecutó en el clúster principal o en un clúster de escalado de simultaneidad. Los valores posibles son los siguientes:<br>0: se ejecutó en el clúster principal<br>1: se ejecutó en un clúster de escalado de simultaneidad<br>Un número mayor que 1: se ejecutó en el clúster principal |

Consultas de ejemplo

La siguiente consulta enumera las cinco consultas más recientes.

```
select query, trim(querytxt) as sqlquery
from stl_query
order by query desc limit 5;

query |                                  sqlquery
------+--------------------------------------------------
```

```
129 | select query, trim(querytxt) from stl_query order by query;
128 | select node from stv_disk_read_speeds;
127 | select system_status from stv_gui_status
126 | select * from systable_topology order by slice
125 | load global dict registry
(5 rows)
```

La siguiente consulta devuelve el tiempo transcurrido en orden descendente para las consultas ejecutadas el 15 de febrero de 2013.

```
select query, datediff(seconds, starttime, endtime),
trim(querytxt) as sqlquery
from stl_query
where starttime >= '2013-02-15 00:00' and endtime < '2013-02-16 00:00'
order by date_diff desc;

 query | date_diff |  sqlquery
-------+-----------+-----------------------------------------
 55    |       119 | padb_fetch_sample: select count(*) from category
121    |         9 | select * from svl_query_summary;
181    |         6 | select * from svl_query_summary where query in(179,178);
172    |         5 | select * from svl_query_summary where query=148;
...
(189 rows)
```

La siguiente consulta muestra el tiempo de cola y el tiempo de ejecución de las consultas. Las consultas con `concurrency_scaling_status = 1` se ejecutaron en un clúster de escalado de simultaneidad. Todas las demás consultas se ejecutaron en el clúster principal.

```
SELECT w.service_class AS queue
     , q.concurrency_scaling_status
     , COUNT( * ) AS queries
     , SUM( q.aborted )  AS aborted
     , SUM( ROUND( total_queue_time::NUMERIC / 1000000,2 ) ) AS queue_secs
     , SUM( ROUND( total_exec_time::NUMERIC / 1000000,2 ) )  AS exec_secs
FROM stl_query q
     JOIN stl_wlm_query w
          USING (userid,query)
WHERE q.userid > 1
  AND service_class > 5
  AND q.starttime > '2019-03-01 16:38:00'
  AND q.endtime   < '2019-03-01 17:40:00'
```

```
GROUP BY 1,2
ORDER BY 1,2;
```

## STL_QUERY_METRICS

Presenta información acerca de las métricas, como la cantidad de filas procesadas, el uso de la CPU, la entrada/salida y el uso del disco para las consultas que se terminaron de ejecutar en colas de consultas (clases de servicios) definidas por el usuario. Para ver las métricas de las consultas activas que están actualmente en ejecución, consulte la vista de sistema STV_QUERY_METRICS.

Las métricas de consultas se muestran en intervalos de un segundo. Por este motivo, diferentes ejecuciones de la misma consulta pueden devolver horas que presentan una pequeña diferencia. Además, es posible que los segmentos de las consultas que se ejecutan en menos de un segundo no se registren.

STL_QUERY_METRICS hace un seguimiento y agrega métricas en el nivel consulta, segmento y paso. Para obtener más información acerca de los segmentos y los pasos de las consultas, consulte Flujo de trabajo de planificación y ejecución de consultas. Distintas métricas (como `max_rows`, `cpu_time`, etc.) se suman en todos los sectores de un nodo. Para obtener más información acerca de los sectores de nodos, consulte Arquitectura del sistema de almacenamiento de datos.

Para determinar en qué nivel de fila se informa sobre las métricas, examine las columnas `segment` y `step_type`.

- Si tanto `segment` como `step_type` son igual a `-1`, la fila se informa sobre las métricas en el nivel consulta.
- Si `segment` no es `-1` y `step_type` es `-1`, la fila se informa sobre las métricas en el nivel segmento.
- Si tanto `segment` como `step_type` no son `-1`, la fila se informa sobre las métricas en el nivel paso.

La vista SVL_QUERY_METRICS y la vista SVL_QUERY_METRICS_SUMMARY agrupan los datos en esta vista y presentan la información de una manera más accesible.

STL_QUERY_METRICS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que ejecutó la consulta que generó la entrada. |
| service_class | entero | ID para la clase de servicio. Las colas de consultas se definen en la configuración de WLM. Las métricas solo se informan para las colas definidas por el usuario. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| segment | entero | Número del segmento. Una consulta consta de distintos segmentos y cada segmento consta de uno o más pasos. Los segmentos de una consulta se ejecutan en paralelo. Cada segmento se ejecuta en un solo proceso. Si el valor del segmento es -1, los valores métricos del segmento se acumulan hasta el nivel consulta. |
| step_type | entero | Tipo de paso que se ejecutó. Para obtener una descripción de los tipos de pasos, consulte Tipos de pasos. |
| starttime | marca de tiempo | Hora en UTC en la que la consulta comenzó a ejecutarse, con 6 dígitos de precisión para los segundos fraccionados. Por ejemplo: `2009-06-12 11:29:19.131358` . |
| slices | entero | Cantidad de sectores para el clúster. |
| max_rows | bigint | Cantidad máxima de filas de salida para un paso, acumulada s en todos los sectores. |
| rows | bigint | Cantidad de filas procesadas por un paso. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| max_cpu_time | bigint | Tiempo de CPU máximo utilizado, en microsegundos. En el nivel segmento, el tiempo de CPU máximo utilizado por el segmento en todos los sectores. En el nivel consulta, el tiempo de CPU máximo utilizado por cualquier segmento de la consulta. |
| cpu_time | bigint | Tiempo de CPU utilizado, en microsegundos. En el nivel segmento, el tiempo de CPU total para el segmento en todos los sectores. En el nivel consulta, la suma de los tiempos de CPU para la consulta en todos los sectores y segmentos. |
| max_block s_read | bigint | Cantidad máxima de bloques de 1 MB leídos por el segmento, acumulados en todos los sectores. En el nivel segmento, la cantidad máxima de bloques de 1 MB leídos para el segmento en todos los sectores. En el nivel consulta, la cantidad máxima de bloques de 1 MB leídos por cualquier segmento de la consulta. |
| blocks_read | bigint | Cantidad de bloques de 1 MB leídos por la consulta o por el segmento. |
| max_run_time | bigint | El tiempo máximo transcurrido para un segmento, en microsegundos. En el nivel segmento, el tiempo máximo de ejecución para el segmento en todos los sectores. En el nivel consulta, el tiempo máximo de ejecución para cualquier segmento de la consulta. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| run_time | bigint | Tiempo de ejecución total, sumado en todos los sectores. El tiempo de ejecución no incluye el tiempo de espera.<br><br>En el nivel segmento, el tiempo de ejecución para el segmento, sumado en todos los sectores. En el nivel consulta, el tiempo de ejecución para la consulta, sumado en todos los sectores y segmentos. Como este valor es una suma, el tiempo de ejecución no está relacionado con el tiempo de ejecución de la consulta. |
| max_blocks_to_disk | bigint | La cantidad máxima de espacio en disco utilizada para grabar los resultados intermedios, en bloques de MB. En el nivel segmento, la cantidad máxima de espacio en disco utilizada por el segmento en todos los sectores. En el nivel consulta, la cantidad máxima de espacio en disco utilizada por cualquier segmento de la consulta. |
| blocks_to_disk | bigint | La cantidad de espacio en disco utilizada por una consulta o un segmento para grabar los resultados intermedios, en bloques de MB. |
| paso | entero | Paso de la consulta que se ejecutó. |
| max_query_scan_size | bigint | El tamaño máximo de los datos examinados por una consulta, en MB. En el nivel segmento, el tamaño máximo de datos examinados por el segmento en todos los sectores. En el nivel consulta, el tamaño máximo de datos examinados por cualquier segmento de la consulta. |
| query_scan_size | bigint | El tamaño de los datos examinados por una consulta, en MB. |
| query_priority | entero | La prioridad de la consulta. Los valores posibles son `-1`, `0`, `1`, `2`, `3` y `4`, donde `-1` significa que no se admite la prioridad de la consulta. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| query_queue_time | bigint | La cantidad de tiempo en microsegundos que la consulta estuvo en cola. |
| service_class_name | character (64) | El nombre de la clase de servicio. |

Consulta de ejemplo

Para encontrar consultas con un tiempo alto de CPU (más de 1 000 segundos), ejecute la siguiente consulta.

```
Select query, cpu_time / 1000000 as cpu_seconds
from stl_query_metrics where segment = -1 and cpu_time > 1000000000
order by cpu_time;

query | cpu_seconds
------+------------
25775 |        9540
```

Para encontrar consultas activas con una combinación de bucle anidado que devolvieron más de un millón de filas, ejecute la siguiente consulta.

```
select query, rows
from stl_query_metrics
where step_type = 15 and rows > 1000000
order by rows;

query | rows
------+-----------
25775 | 2621562702
```

Para encontrar consultas activas que se ejecutaron por más de 60 segundos y que usaron menos de 10 segundos de tiempo de CPU, ejecute la siguiente consulta.

```
select query, run_time/1000000 as run_time_seconds
from stl_query_metrics
```

```
where segment = -1 and run_time > 60000000 and cpu_time < 10000000;

query | run_time_seconds
------+-----------------
25775 |              114
```

## STL_QUERYTEXT

Captura el texto de la consulta para los comandos SQL.

Consulte la vista STL_QUERYTEXT para capturar el SQL que se registró para las siguientes instrucciones:

- SELECT, SELECT INTO
- INSERT, UPDATE, DELETE
- COPY
- UNLOAD
- Consultas generadas al ejecutar VACUUM y ANALYZE
- CREATE TABLE AS (CTAS)

Para consultar la actividad de estas instrucciones durante un periodo determinado, combine las vistas STL_QUERYTEXT y STL_QUERY.

> **Note**
>
> Las vistas STL_QUERY y STL_QUERYTEXT solo presentan información relacionada con consultas, no sobre otras utilidades y comandos de DDL. Para obtener una lista con información sobre todas las instrucciones ejecutadas por Amazon Redshift, también puede consultar las vistas STL_DDLTEXT y STL_UTILITYTEXT. Para obtener una lista completa de todas las instrucciones ejecutadas por Amazon Redshift, puede consultar la vista SVL_STATEMENTTEXT.

Consulte también STL_DDLTEXT, STL_UTILITYTEXT y SVL_STATEMENTTEXT.

STL_QUERYTEXT es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_QUERY_TEXT. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| xid | bigint | ID de transacción. |
| pid | entero | ID de proceso. Por lo general, todas las consultas en una sesión se ejecutan en el mismo proceso, por lo que este valor suele permanecer constante si ejecuta una serie de consultas en la misma sesión. Después de determina dos eventos internos, Amazon Redshift podría reiniciar una sesión activa y asignar un PID nuevo. Para obtener más información, consulte STL_RESTARTED_SESSIONS. Puede utilizar esta columna para combinar la vista STL_ERROR. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| sequence | entero | Cuando una única instrucción tiene más de 200 caracteres, se registran filas adicionales para esa instrucción. Secuencia 0 es la primera fila, 1 es la segunda y, así, sucesivamente. |
| texto | character(200) | Texto SQL, en incrementos de 200 caracteres. El campo puede contener caracteres especiales como una barra inversa(\\) y una nueva línea (\n). |

### Consultas de ejemplo

Puede utilizar la función PG_BACKEND_PID() para recuperar información de la sesión actual. Por ejemplo, la siguiente consulta devuelve el ID de consulta y una parte de su texto para aquellas consultas que se hayan completado en la sesión actual.

```
select query, substring(text,1,60)
from stl_querytext
where pid = pg_backend_pid()
order by query desc;

 query |                          substring
-------+------------------------------------------------------------
 28262 | select query, substring(text,1,80) from stl_querytext where
 28252 | select query, substring(path,0,80) as path from stl_unload_l
 28248 | copy category from 's3://dw-tickit/manifest/category/1030_ma
 28247 | Count rows in target table
 28245 | unload ('select * from category') to 's3://dw-tickit/manifes
 28240 | select query, substring(text,1,40) from stl_querytext where
(6 rows)
```

### Reconstruyendo el SQL almacenado

Para reconstruir el SQL almacenado en la columna `text` de STL_QUERYTEXT, ejecute una instrucción SELECT para crear SQL de 1 o más partes en la columna `text`. Antes de ejecutar el SQL reconstruido, reemplace cualquier caracter especial (\n) con una nueva línea. El resultado de la siguiente instrucción SELECT es filas de SQL reconstruido en el campo `query_statement`.

```
SELECT query, LISTAGG(CASE WHEN LEN(RTRIM(text)) = 0 THEN text ELSE RTRIM(text) END)
 WITHIN GROUP (ORDER BY sequence) as query_statement, COUNT(*) as row_count
FROM stl_querytext GROUP BY query ORDER BY query desc;
```

Por ejemplo, la siguiente búsqueda selecciona 3 columnas. La búsqueda en si misma es más larga de 200 caracteres y se guarda en varias partes en STL_QUERYTEXT.

```
select
1 AS a01234567890123456789012345678901234567890123456789012345678901234567890,
2 AS b01234567890123456789012345678901234567890123456789012345678901234567890,
3 AS b012345678901234567890123456789012345678901234
FROM stl_querytext;
```

En este ejemplo, la búsqueda se guarda en 2 partes (filas) en la columna `text` de STL_QUERYTEXT.

```
select query, sequence, text
from stl_querytext where query=pg_last_query_id() order by query desc, sequence limit
 10;
```

```
query  | sequence |
                        text

-------+----------
+-----------------------------------------------------------------------------------------------------
    45 |        0 | select\n1 AS
 a012345678901234567890123456789012345678901234567890123456789012345678901234567890,\n2 AS
 b012345678901234567890123456789012345678901234567890123456789012345678901234567890,\n3 AS
 b012345678901234567890123456789012345678901234
    45 |        1 | \nFROM stl_querytext;
```

Para reconstruir el SQL almacenado en STL_QUERYTEXT, ejecute el siguiente SQL.

```
select LISTAGG(CASE WHEN LEN(RTRIM(text)) = 0 THEN text ELSE RTRIM(text) END, '')
 within group (order by sequence) AS text
from stl_querytext where query=pg_last_query_id();
```

Para usar el SQL reconstruido resultante en su cliente, reemplace cualquier (\n) caracter especial con una nueva línea.

```
                    text

-----------------------------------------------------------------------------------------------------
 select\n1 AS a012345678901234567890123456789012345678901234567890123456789012345678901234567890,
\n2 AS b012345678901234567890123456789012345678901234567890123456789012345678901234567890,\n3 AS
 b012345678901234567890123456789012345678901234\nFROM stl_querytext;
```

## STL_REPLACEMENTS

Muestra un registro de los momentos en que se reemplazaron caracteres UTF-8 no válidos por el comando COPY con la opción ACCEPTINVCHARS. Se agrega una entrada de registro

a STL_REPLACEMENTS por cada una de las 100 primeras filas de cada sector del nodo que necesitaron al menos un reemplazo.

STL_REPLACEMENTS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_NESTLOOP solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_COPY_REPLACEMENTS. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| slice | entero | Número del sector del nodo en el que se produjo el reemplazo . |
| tbl | entero | ID de la tabla. |
| starttime | marca de tiempo | La hora de inicio en UTC del comando COPY. |
| sesión | entero | ID de la sesión que realiza el comando COPY. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| filename | character (256) | Ruta completa hacia el archivo de entrada para el comando COPY. |
| line_number | bigint | Número de línea en el archivo de datos de entrada que tiene un carácter UTF-8 no válido. Un -1 indica que el número de línea no está disponible, por ejemplo, al copiar desde un archivo de datos con formato de columna. |
| colname | character (127) | Primer campo que tiene un carácter UTF-8 no válido. |
| raw_line | character (1024) | Datos de carga sin procesar que tienen un carácter UTF-8 no válido. |

Consultas de ejemplo

El siguiente ejemplo devuelve los reemplazos de la operación COPY más reciente.

```
select query, session, filename, line_number, colname
from stl_replacements
where query = pg_last_copy_id();

 query | session |   filename                                  | line_number | colname
 ------+---------+---------------------------------------------+-------------+--------
    96 |    6314 | s3://DOC-EXAMPLE-BUCKET/allusers_pipe.txt   |         251 | city
    96 |    6314 | s3://DOC-EXAMPLE-BUCKET/allusers_pipe.txt   |         317 | city
    96 |    6314 | s3://DOC-EXAMPLE-BUCKET/allusers_pipe.txt   |         569 | city
    96 |    6314 | s3://DOC-EXAMPLE-BUCKET/allusers_pipe.txt   |         623 | city
    96 |    6314 | s3://DOC-EXAMPLE-BUCKET/allusers_pipe.txt   |         694 | city
...
```

## STL_RESTARTED_SESSIONS

Para mantener una disponibilidad constante después de determinados eventos internos, Amazon Redshift podría reiniciar una sesión activa con un ID de proceso (PID) nuevo. Cuando Amazon Redshift reinicia una sesión, STL_RESTARTED_SESSIONS registra el PID nuevo y el PID anterior.

Para obtener más información, vea los ejemplos a continuación en esta sección.

STL_RESTARTED_SESSIONS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_SESSION_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| currenttime | marca de tiempo | Hora del evento. |
| dbname | character (50) | Nombre de la base de datos asociada a la sesión. |
| newpid | entero | ID de proceso para la sesión reiniciada. |
| oldpid | entero | ID de proceso para la sesión original. |
| nombre de usuario | character (50) | Nombre del usuario asociado a la sesión. |
| remotehost | character (45) | Nombre o dirección IP del host remoto. |
| remoteport | character (32) | Número del puerto del host remoto. |
| parkedtime | marca de tiempo | Esta información es solo para uso interno. |
| session_vars | character (2000) | Esta información es solo para uso interno. |

### Consultas de ejemplo

En el siguiente ejemplo, se combina STL_RESTARTED_SESSIONS con STL_SESSIONS para mostrar los nombres de usuario para las sesiones que se reiniciaron.

```
select process, stl_restarted_sessions.newpid, user_name
from stl_sessions
inner join stl_restarted_sessions on stl_sessions.process =
 stl_restarted_sessions.oldpid
order by process;

...
```

## STL_RETURN

Presenta detalles sobre los pasos de devolución de las consultas. Un paso de devolución devuelve los resultados de las consultas completadas en los nodos de informática al nodo principal. El nodo principal luego fusiona los datos y devuelve los resultados al cliente que los solicita. En el caso de las consultas completadas en el nodo principal, un paso de devolución devuelve los resultados al cliente.

Una consulta consta de distintos segmentos y cada segmento consta de uno o más pasos. Para obtener más información, consulte Procesamiento de consultas.

STL_RETURN es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_RETURN solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| slice | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |
| segment | entero | Número que identifica el segmento de la consulta. |
| paso | entero | Paso de la consulta que se ejecutó. |
| starttime | marca de tiempo | Hora en UTC en que comenzó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| endtime | marca de tiempo | Hora en UTC en la que finalizó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| tasknum | entero | Es el número del proceso de tarea de la consulta que se asignó para ejecutar el paso. |
| rows | bigint | Es la cantidad total de filas que se procesaron. |
| bytes | bigint | Tamaño, en bytes, de todas las filas de resultado del paso. |
| packets | entero | Cantidad total de paquetes enviados por la red. |
| checksum | bigint | Esta información es solo para uso interno. |

Consultas de ejemplo

La siguiente consulta muestra qué pasos de la consulta más reciente se realizaron en cada sector.

```
SELECT query, slice, segment, step, endtime, rows, packets
from stl_return where query = pg_last_query_id();

 query |  slice | segment | step |          endtime           | rows | packets
-------+--------+---------+------+----------------------------+------+---------
     4 |      2 |       3 |    2 | 2013-12-27 01:43:21.469043 |    3 |       0
     4 |      3 |       3 |    2 | 2013-12-27 01:43:21.473321 |    0 |       0
     4 |      0 |       3 |    2 | 2013-12-27 01:43:21.469118 |    2 |       0
     4 |      1 |       3 |    2 | 2013-12-27 01:43:21.474196 |    0 |       0
     4 |      4 |       3 |    2 | 2013-12-27 01:43:21.47704  |    2 |       0
     4 |      5 |       3 |    2 | 2013-12-27 01:43:21.478593 |    0 |       0
     4 |   12811|       4 |    1 | 2013-12-27 01:43:21.480755 |    0 |       0
(7 rows)
```

## STL_S3CLIENT

Registra el tiempo de transferencia y otras métricas de rendimiento.

Utilice la tabla STL_S3CLIENT para encontrar el tiempo que duró la transferencia de datos desde Amazon S3.

STL_S3CLIENT es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| slice | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |
| recordtime | marca de tiempo | Hora en la que se graba el registro. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| pid | entero | ID de proceso. Todas las consultas en una sesión se ejecutan en el mismo proceso, por lo que este valor permanece constante si ejecuta una serie de consultas en la misma sesión. |
| http_method | character (64) | Nombre del método HTTP que corresponde a la solicitud de Amazon S3. |
| bucket | character (64) | Nombre del bucket de S3. |
| clave | character (256) | Clave que corresponde al objeto de Amazon S3. |
| transfer_size | bigint | Cantidad de bytes transferidos. |
| data_size | bigint | Cantidad de bytes de datos. Este valor es el mismo que el de transfer_size para los datos no comprimidos. Si se utilizó una compresión, este es el tamaño de los datos no comprimidos. |
| start_time | bigint | Hora en la que comenzó la transferencia (en microsegundos desde el 1 de enero del 2000). |
| end_time | bigint | Hora en la que terminó la transferencia (en microsegundos desde 1 de enero de 2000). |
| transfer_time | bigint | Tiempo que demoró la transferencia (en microsegundos). |
| compression_time | bigint | Parte del tiempo de transferencia que tardó la descompresión de datos (en microsegundos). |
| connect_time | bigint | Tiempo desde el inicio hasta que se completó la conexión al servidor remoto (en microsegundos). |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| app_connect_time | bigint | Tiempo desde el inicio hasta que se completó la conexión SSL o el protocolo de enlace con el host remoto (en microsegundos). |
| retries | bigint | Cantidad de veces que se reintentó hacer la transferencia. |
| request_id | char(32) | ID de la solicitud del encabezado de respuesta de HTTP de Amazon S3 |
| extended_ request_id | char(128) | ID de la solicitud ampliada de la respuesta del encabezado de HTTP de Amazon S3 (x-amz-id-2). |
| ip_address | char(64) | Dirección IP del servidor (ip V4 o V6). |
| is_partial | entero | Valor que, si es true (1), indica que el archivo de entrada se divide en rangos durante una operación COPY. Si este valor es false (0), el archivo de entrada no se divide. |
| start_offset | bigint | Valor que, si el archivo de entrada se divide durante una operación COPY, indica el valor de desplazamiento de la división (en bytes). Si el archivo no se divide, este valor es 0. |

### Consulta de ejemplo

La siguiente consulta devuelve el tiempo que tardó la carga de archivos con el comando COPY.

```
select slice, key, transfer_time
from stl_s3client
where query = pg_last_copy_id();
```

### Resultado

```
 slice |   key                        | transfer_time
------+------------------------------+---------------
     0 | listing10M0003_part_00       |    16626716
     1 | listing10M0001_part_00       |    12894494
```

```
     2 | listing10M0002_part_00      |    14320978
     3 | listing10M0000_part_00      |    11293439
  3371 | prefix=listing10M;marker=   |       99395
```

La siguiente búsqueda convierte el `start_time` y `end_time` en una marca de hora.

```
select userid,query,slice,pid,recordtime,start_time,end_time,
'2000-01-01'::timestamp + (start_time/1000000.0)* interval '1 second' as start_ts,
'2000-01-01'::timestamp + (end_time/1000000.0)* interval '1 second' as end_ts
from stl_s3client where query> -1 limit 5;
```

```
 userid | query | slice |  pid  |         recordtime         |   start_time    |
 end_time     |          start_ts          |           end_ts
--------+-------+-------+-------+----------------------------+-----------------
+-----------------+----------------------------+----------------------------
      0 |     0 |     0 | 23449 | 2019-07-14 16:27:17.207839 | 616436837154256 |
 616436837207838 | 2019-07-14 16:27:17.154256 | 2019-07-14 16:27:17.207838
      0 |     0 |     0 | 23449 | 2019-07-14 16:27:17.252521 | 616436837208208 |
 616436837252520 | 2019-07-14 16:27:17.208208 | 2019-07-14 16:27:17.25252
      0 |     0 |     0 | 23449 | 2019-07-14 16:27:17.284376 | 616436837208460 |
 616436837284374 | 2019-07-14 16:27:17.20846  | 2019-07-14 16:27:17.284374
      0 |     0 |     0 | 23449 | 2019-07-14 16:27:17.285307 | 616436837208980 |
 616436837285306 | 2019-07-14 16:27:17.20898  | 2019-07-14 16:27:17.285306
      0 |     0 |     0 | 23449 | 2019-07-14 16:27:17.353853 | 616436837302216 |
 616436837353851 | 2019-07-14 16:27:17.302216 | 2019-07-14 16:27:17.353851
```

## STL_S3CLIENT_ERROR

Registra los errores que surgieron en un sector mientras se cargaba un archivo desde Amazon S3.

Utilice STL_S3CLIENT_ERROR para encontrar detalles acerca de los errores que surgieron mientras se transferían datos desde Amazon S3 como parte de un comando COPY.

STL_S3CLIENT_ERROR es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. El ID de consulta -1 es para uso interno. |
| sliceid | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |
| recordtime | marca de tiempo | Hora en la que se graba el registro. |
| pid | entero | ID de proceso. Todas las consultas en una sesión se ejecutan en el mismo proceso, por lo que este valor permanece constante si ejecuta una serie de consultas en la misma sesión. |
| http_method | character (64) | Nombre del método HTTP que corresponde a la solicitud de Amazon S3. |
| bucket | character (64) | Nombre de bucket de Amazon S3. |
| clave | character (256) | Clave que corresponde al objeto de Amazon S3. |
| error | character (1024) | Mensaje de error. |
| is_partial | entero | Valor que, si es true (1), indica que el archivo de entrada se divide en rangos durante una operación COPY. Si este valor es false (0), el archivo de entrada no se divide. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| start_offset | bigint | Valor que, si el archivo de entrada se divide durante una operación COPY, indica el valor de desplazamiento de la división (en bytes). Si el archivo no se divide, este valor es 0. |

Notas de uso

Si ve errores múltiples con "Connection timed out", es posible que tenga un problema de redes. Si está usando la función Enhanced VPC Routing, verifique que tiene una ruta de acceso a la red válida entre la VPC de su clúster y sus recursos de datos. Para obtener más información, consulte Enrutamiento de VPC mejorada de Amazon Redshift.

Consulta de ejemplo

La siguiente consulta devuelve los errores de los comandos COPY completados durante la sesión actual.

```
select query, sliceid, substring(key from 1 for 20) as file,
substring(error from 1 for 35) as error
from stl_s3client_error
where pid = pg_backend_pid()
order by query desc;
```

Resultado

```
 query  | sliceid |     file           |             error

--------+---------+--------------------+-----------------------------------
 362228 |      12 | part.tbl.25.159.gz | transfer closed with 1947655 bytes
 362228 |      24 | part.tbl.15.577.gz | transfer closed with 1881910 bytes
 362228 |       7 | part.tbl.22.600.gz | transfer closed with 700143 bytes r
 362228 |      22 | part.tbl.3.34.gz   | transfer closed with 2334528 bytes
 362228 |      11 | part.tbl.30.274.gz | transfer closed with 699031 bytes r
 362228 |      30 | part.tbl.5.509.gz  | Unknown SSL protocol error in conne
 361999 |      10 | part.tbl.23.305.gz | transfer closed with 698959 bytes r
 361999 |      19 | part.tbl.26.582.gz | transfer closed with 1881458 bytes
 361999 |       4 | part.tbl.15.629.gz | transfer closed with 2275907 bytes
```

```
 361999 |      20 | part.tbl.6.456.gz  | transfer closed with 692162 bytes r
(10 rows)
```

## STL_SAVE

Presenta detalles sobre los pasos para guardar de las consultas. Un paso guardar guarda la secuencia de entrada en una tabla transitoria. Una tabla transitoria es una tabla temporal que almacena resultados intermedios durante la ejecución de una consulta.

Una consulta consta de distintos segmentos y cada segmento consta de uno o más pasos. Para obtener más información, consulte Procesamiento de consultas.

STL_SAVE es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_SAVE solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| slice | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| segment | entero | Número que identifica el segmento de la consulta. |
| paso | entero | Paso de la consulta que se ejecutó. |
| starttime | marca de tiempo | Hora en UTC en que comenzó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| endtime | marca de tiempo | Hora en UTC en la que finalizó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| tasknum | entero | Es el número del proceso de tarea de la consulta que se asignó para ejecutar el paso. |
| rows | bigint | Es la cantidad total de filas que se procesaron. |
| bytes | bigint | Tamaño, en bytes, de todas las filas de resultado del paso. |
| tbl | entero | ID de la tabla transitoria materializada. |
| is_diskbased | character (1) | Indica si este paso de la consulta se ejecutó como una operación basada en el disco: true (`t`) (verdadero) o false (`f`) (falso). |
| workmem | bigint | Cantidad de bytes de memoria funcional asignada a este paso. |

Consultas de ejemplo

La siguiente consulta muestra qué pasos de guardado de la consulta más reciente se realizaron en cada sector.

```
select query, slice, segment, step, tasknum, rows,  tbl
from stl_save where query = pg_last_query_id();

 query | slice | segment | step | tasknum | rows | tbl
-------+-------+---------+------+---------+------+-----
 52236 |     3 |       0 |    2 |      21 |    0 | 239
```

```
 52236 |     2 |       0 |    2 |      20 |    0 | 239
 52236 |     2 |       2 |    2 |      20 |    0 | 239
 52236 |     3 |       2 |    2 |      21 |    0 | 239
 52236 |     1 |       0 |    2 |      21 |    0 | 239
 52236 |     0 |       0 |    2 |      20 |    0 | 239
 52236 |     0 |       2 |    2 |      20 |    0 | 239
 52236 |     1 |       2 |    2 |      21 |    0 | 239
(8 rows)
```

## STL_SCAN

Analiza los pasos de exámenes de tablas de las consultas. El número de paso para las filas de esta tabla siempre es 0 porque un examen es el primer paso de un segmento.

STL_SCAN es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_SCAN solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| slice | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |
| segment | entero | Número que identifica el segmento de la consulta. |
| paso | entero | Paso de la consulta que se ejecutó. |
| starttime | marca de tiempo | Hora en UTC en que comenzó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| endtime | marca de tiempo | Hora en UTC en la que finalizó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| tasknum | entero | Es el número del proceso de tarea de la consulta que se asignó para ejecutar el paso. |
| rows | bigint | Es la cantidad total de filas que se procesaron. |
| bytes | bigint | Tamaño, en bytes, de todas las filas de resultado del paso. |
| fetches | bigint | Esta información es solo para uso interno. |
| type | entero | ID del tipo de examen. Para obtener una lista de valores válidos, consulte la siguiente tabla. |
| tbl | entero | ID de la tabla. |
| is_rrscan | character (1) | Si es true (t), indica que en el paso se utilizó un examen de rango restringido. |
| is_delayed_scan | character (1) | Esta información es solo para uso interno. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| rows_pre_filter | bigint | Para exámenes de tablas permanentes, la cantidad total de filas emitidas antes de filtrar las filas marcadas para eliminación (filas fantasma) y antes de aplicar los filtros de consulta definidos por el usuario. |
| rows_pre_user_filter | bigint | Para los exámenes de tablas permanentes, la cantidad total de filas procesadas después de filtrar las marcadas para eliminación (filas fantasma), pero antes de aplicar los filtros de consulta definidos por el usuario. |
| perm_table_name | character (136) | Para los exámenes de tablas permanentes, el nombre de la tabla examinada. |
| is_rlf_scan | character (1) | Si es true (t), indica que en el paso se utilizó el filtrado en el nivel de fila. |
| is_rlf_scan_reason | entero | Esta información es solo para uso interno. |
| num_em_blocks | entero | Esta información es solo para uso interno. |
| checksum | bigint | Esta información es solo para uso interno. |
| runtime_filtering | character (1) | Si es true (t), indica que se aplican filtros en tiempo de ejecución. |
| scan_region | entero | Esta información es solo para uso interno. |
| num_sortkey_as_predicate | entero | Esta información es solo para uso interno. |
| row_fetcher_state | entero | Esta información es solo para uso interno. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| consumed_scan_ranges | bigint | Esta información es solo para uso interno. |
| work_stealing_reason | bigint | Esta información es solo para uso interno. |
| is_vectorized_scan | character (1) | Esta información es solo para uso interno. |
| is_vectorized_scan_reason | entero | Esta información es solo para uso interno. |
| row_fetcher_reason | bigint | Esta información es solo para uso interno. |
| topology_signature | bigint | Esta información es solo para uso interno. |
| use_tpm_partition | character (1) | Esta información es solo para uso interno. |
| is_rrscan_expr | character (1) | Esta información es solo para uso interno. |
| scanned_mega_value | character (1) | Esta información es solo para uso interno. Esta información muestra si el paso de escaneo dado ha escaneado un valor grande. Un valor grande se almacenará en varios bloques. El tamaño del bloque es de 1 MB de forma predeterminada, un valor grande es superior a 1 MB en la configuración predeterminada. |

Tipos de exámenes

| ID del tipo | Descripción |
|---|---|
| 1 | Datos de la red. |
| 2 | Tablas permanentes de usuarios en la memoria compartida comprimida. |
| 3 | Tablas transitorias en filas. |
| 21 | Carga de archivos desde Amazon S3. |
| 22 | Carga de tablas desde Amazon DynamoDB. |
| 23 | Carga de datos desde una conexión SSH remota. |
| 24 | Carga de datos desde un clúster remoto (región ordenada). Se utiliza para el cambio de tamaño. |
| 25 | Carga de datos desde un clúster remoto (región desordenada). Se utiliza para el cambio de tamaño. |
| 28 | Lea los datos de una vista de serie temporal con UNION ALL en varias tablas. |
| 29 | Lectura de datos desde tablas externas de Amazon S3. |
| 30 | Lectura de la información de partición de una tabla externa de Amazon S3. |
| 33 | Lectura de datos desde una tabla de Postgres remota. |
| 36 | Lectura de datos desde una tabla de MySQL remota. |
| 37 | lectura de datos de una transmisión remota de Kinesis. |

Notas de uso

Idealmente, `rows` debe tener un valor relativamente cercano a `rows_pre_filter`. Una gran diferencia entre `rows` y `rows_pre_filter` indica que el motor de ejecución está examinando filas que luego se descartan, lo cual es ineficaz. La diferencia entre `rows_pre_filter` y `rows_pre_user_filter` es la cantidad de filas fantasmas en el examen. Ejecute una función

VACUUM para eliminar las filas marcadas para ser eliminadas. La diferencia entre `rows` y `rows_pre_user_filter` es la cantidad de filas filtradas por la consulta. Si se descartan muchas filas por el filtro del usuario, revise su elección de columna de ordenación o si el descarte se debe a una región desordenada de gran tamaño ejecute una limpieza.

Consultas de ejemplo

En el siguiente ejemplo, se muestra que `rows_pre_filter` es mayor que `rows_pre_user_filter` porque la tabla ha eliminado filas que no fueron limpiadas (filas fantasma).

```
SELECT query, slice, segment,step,rows, rows_pre_filter, rows_pre_user_filter
from stl_scan where query = pg_last_query_id();

 query |  slice | segment | step | rows  | rows_pre_filter | rows_pre_user_filter
-------+--------+---------+------+-------+-----------------+----------------------
 42915 |      0 |       0 |    0 | 43159 |           86318 |                43159
 42915 |      0 |       1 |    0 |     1 |               0 |                    0
 42915 |      1 |       0 |    0 | 43091 |           86182 |                43091
 42915 |      1 |       1 |    0 |     1 |               0 |                    0
 42915 |      2 |       0 |    0 | 42778 |           85556 |                42778
 42915 |      2 |       1 |    0 |     1 |               0 |                    0
 42915 |      3 |       0 |    0 | 43428 |           86856 |                43428
 42915 |      3 |       1 |    0 |     1 |               0 |                    0
 42915 |  10000 |       2 |    0 |     4 |               0 |                    0
(9 rows)
```

## STL_SCHEMA_QUOTA_VIOLATIONS

Registra la incidencia, marca de tiempo, XID y otros datos útiles cuando se excede una cuota de esquema.

STL_SCHEMA_QUOTA_VIOLATIONS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_SCHEMA_QUOTA_VIOLATIONS. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| ownerid | entero | El ID del propietario del esquema. |
| xid | bigint | El ID de la transacción asociado a la instrucción. |
| pid | entero | El ID del proceso asociado a la instrucción. |
| userid | entero | ID del usuario que generó la entrada. |
| schema_id | entero | El espacio de nombres o el ID de esquema. |
| schema_name | carácter (128) | El espacio de nombres o el nombre del esquema. |
| cuota | entero | La cantidad de espacio en disco (en MB) que puede utilizar el esquema. |
| disk_usage | entero | El espacio en disco (en MB) que utiliza actualmente el esquema. |
| disk_usage_pct | double precision | El porcentaje de espacio en disco que utiliza actualmente el esquema fuera de la cuota configurada. |
| marca de tiempo | TIMESTAMP sin zona horaria | La hora en que se produjo la infracción. |

Consultas de ejemplo

La siguiente consulta muestra el resultado de la infracción de cuota:

```
SELECT userid, TRIM(SCHEMA_NAME) "schema_name", quota, disk_usage, disk_usage_pct,
 timestamp FROM
stl_schema_quota_violations WHERE SCHEMA_NAME = 'sales_schema' ORDER BY timestamp DESC;
```

Esta consulta devuelve el siguiente resultado de ejemplo para el esquema especificado:

```
userid | schema_name  | quota | disk_usage | disk_usage_pct |timestamp
-------+--------------+-------+------------+----------------
+----------------------------
104    | sales_schema | 2048  | 2798       |  136.62        | 2020-04-20
 20:09:25.494723
(1 row)
```

## STL_SESSIONS

Devuelve información acerca del historial de sesiones del usuario.

STL_SESSIONS es diferente de STV_SESSIONS en cuanto a que STL_SESSIONS tiene el historial de sesiones, mientras que STV_SESSIONS tiene las sesiones activas actuales.

STL_SESSIONS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_SESSION_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| starttime | marca de tiempo | Hora en UTC en que comenzó la sesión. |
| endtime | marca de tiempo | Hora en UTC en que finalizó la sesión. |
| procesar | entero | ID de proceso para la sesión. |
| user_name | character (50) | Nombre de usuario asociado a la sesión. |
| db_name | character (50) | Nombre de la base de datos asociada a la sesión. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| timeout_sec | int | El tiempo máximo en segundos durante el cual una sesión permanece inactiva o en reposo antes de que se cierre. El 0 indica que no se establece tiempo de espera. |
| timed_out | int | Un valor que indica el motivo por el que se ha terminado la conexión. Puede tener los siguientes valores:<br>• 0: la conexión se ha terminado debido a un error desconocido.<br>• 1: se ha agotado el tiempo de espera de la conexión.<br>• 2: el cliente ha terminado la conexión.<br>• 3: un error interno del backend de Amazon Redshift ha terminado la conexión. |

Consultas de ejemplo

Para ver el historial de sesiones para la base de datos TICKIT, escriba la siguiente consulta:

```
select starttime, process, user_name, timeout_sec, timed_out
from stl_sessions
where db_name='tickit' order by starttime;
```

Esta consulta devuelve el siguiente ejemplo de salida:

```
    starttime              | process |  user_name             | timeout_sec | timed_out
---------------------------+---------+------------------------+-------------
+-------------
2008-09-15 09:54:06.746705 |   32358 | dwuser                 | 120         | 1
2008-09-15 09:56:34.30275  |   32744 | dwuser                 | 60          | 1
2008-09-15 11:20:34.694837 |   14906 | dwuser                 | 0           | 0
2008-09-15 11:22:16.749818 |   15148 | dwuser                 | 0           | 0
```

```
2008-09-15 14:32:44.66112  |   14031 | dwuser                | 0           | 0
2008-09-15 14:56:30.22161  |   18380 | dwuser                | 0           | 0
2008-09-15 15:28:32.509354 |   24344 | dwuser                | 0           | 0
2008-09-15 16:01:00.557326 |   30153 | dwuser                | 120         | 1
2008-09-15 17:28:21.419858 |   12805 | dwuser                | 0           | 0
2008-09-15 20:58:37.601937 |   14951 | dwuser                | 60          | 1
2008-09-16 11:12:30.960564 |   27437 | dwuser                | 60          | 1
2008-09-16 14:11:37.639092 |   23790 | dwuser                | 3600        | 1
2008-09-16 15:13:46.02195  |    1355 | dwuser                | 120         | 1
2008-09-16 15:22:36.515106 |    2878 | dwuser                | 120         | 1
2008-09-16 15:44:39.194579 |    6470 | dwuser                | 120         | 1
2008-09-16 16:50:27.02138  |   17254 | dwuser                | 120         | 1
2008-09-17 12:05:02.157208 |    8439 | dwuser                | 3600        | 0
(17 rows)
```

## STL_SORT

Muestra los pasos de ejecución de ordenación para las consultas; por ejemplo, los pasos que utiliza el procesamiento ORDER BY.

STL_SORT es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_SORT solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| slice | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |
| segment | entero | Número que identifica el segmento de la consulta. |
| paso | entero | Paso de la consulta que se ejecutó. |
| starttime | marca de tiempo | Hora en UTC en que comenzó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| endtime | marca de tiempo | Hora en UTC en la que finalizó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| tasknum | entero | Es el número del proceso de tarea de la consulta que se asignó para ejecutar el paso. |
| rows | bigint | Es la cantidad total de filas que se procesaron. |
| bytes | bigint | Tamaño, en bytes, de todas las filas de resultado del paso. |
| tbl | entero | ID de la tabla. |
| is_diskbased | character (1) | Si es true (t) (verdadero), significa que la consulta se realizó como una operación basada en el disco. Si es false (f) (falso), implica que la consulta se realizó en la memoria. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| workmem | bigint | Cantidad total de bytes en la memoria funcional que se asignaron a este paso. |
| checksum | bigint | Esta información es solo para uso interno. |

### Consultas de ejemplo

El siguiente ejemplo devuelve los resultados de ordenaciones para el sector 0 y el segmento 1.

```
select query, bytes, tbl, is_diskbased, workmem
from stl_sort
where slice=0 and segment=1;
```

```
 query |  bytes  | tbl | is_diskbased |  workmem
-------+---------+-----+--------------+-----------
   567 | 3126968 | 241 | f            | 383385600
   604 |    5292 | 242 | f            | 383385600
   675 |  104776 | 251 | f            | 383385600
   525 | 3126968 | 251 | f            | 383385600
   585 |    5068 | 241 | f            | 383385600
   630 |  204808 | 266 | f            | 383385600
   704 |       0 | 242 | f            |         0
   669 | 4606416 | 241 | f            | 383385600
   696 |  104776 | 241 | f            | 383385600
   651 | 4606416 | 254 | f            | 383385600
   632 |       0 | 256 | f            |         0
   599 |     396 | 241 | f            | 383385600
 86397 |       0 | 242 | f            |         0
   621 |    5292 | 241 | f            | 383385600
 86325 |       0 | 242 | f            |         0
   572 |    5068 | 242 | f            | 383385600
   645 |  204808 | 241 | f            | 383385600
   590 |     396 | 242 | f            | 383385600
(18 rows)
```

## STL_SSHCLIENT_ERROR

Registra todos los errores vistos por el cliente Secure Shell (SSH, Shell seguro).

STL_SSHCLIENT_ERROR es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| slice | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |
| recordtime | marca de tiempo | Hora en la que se registró el error. |
| pid | entero | Proceso que registró el error. |
| ssh_username | character (1024) | Nombre de usuario de SSH. |
| punto de conexión | character (1024) | Punto de conexión de SSH. |
| comando | character (4096) | El comando SSH completo. |
| error | character (1024) | Mensaje de error. |

## STL_STREAM_SEGS

Muestra la relación entre las secuencias y los segmentos simultáneos.

Los flujos en este contexto son flujos de Amazon Redshift. Esta vista del sistema no se refiere a Ingesta de streaming a una vista materializada.

STL_STREAM_SEGS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_STREAM_SEGS solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| secuencia | entero | El conjunto de segmentos simultáneos de una consulta. |
| segment | entero | Número que identifica el segmento de la consulta. |

Consultas de ejemplo

Para ver la relación entre las secuencias y los segmentos simultáneos de la consulta más reciente, escriba la siguiente consulta:

```
select *
from stl_stream_segs
where query = pg_last_query_id();
```

```
 query | stream | segment
-------+--------+---------
    10 |      1 |       2
    10 |      0 |       0
    10 |      2 |       4
    10 |      1 |       3
    10 |      0 |       1
(5 rows)
```

## STL_TR_CONFLICT

Muestra información para identificar y resolver conflictos de transacciones con las tablas de las bases de datos.

Los conflictos de transacciones se producen cuando dos o más usuarios están consultando y modificando filas de datos incluidas en tablas cuyas transacciones no se pueden serializar. La transacción que ejecuta una instrucción que rompería la capacidad de serialización se detiene y se revierte. Cada vez que se produce un conflicto en una transacción, Amazon Redshift escribe una fila de datos en la tabla de sistema STL_TR_CONFLICT con los detalles sobre la transacción cancelada. Para obtener más información, consulte Aislamiento serializable.

STL_TR_CONFLICT solo es visible para los superusuarios. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_TRANSACTION_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| xact_id | bigint | ID de transacción para la transacción revertida. |
| process_id | bigint | Proceso asociado con la transacción revertida. |
| xact_start_ts | marca de tiempo | Hora (UTC) en la que comenzó la transacción. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| abort_time | marca de tiempo | Hora (UTC) en la que se detuvo la transacción. |
| table_id | bigint | ID de la tabla para la tabla en la que se produjo el conflicto. |

Consulta de ejemplo

Para devolver información acerca de los conflictos relacionados con una tabla en particular, ejecute una consulta que especifique el ID de la tabla:

```
select * from stl_tr_conflict where table_id=100234
order by xact_start_ts;

xact_id|process_|      xact_start_ts       |        abort_time        |table_
       |id      |                          |                          |id
-------+--------+--------------------------+--------------------------+------
  1876 |  8551  |2010-03-30 09:19:15.852326|2010-03-30 09:20:17.582499|100234
  1928 | 15034  |2010-03-30 13:20:00.636045|2010-03-30 13:20:47.766817|100234
  1991 | 23753  |2010-04-01 13:05:01.220059|2010-04-01 13:06:06.94098 |100234
  2002 | 23679  |2010-04-01 13:17:05.173473|2010-04-01 13:18:27.898655|100234
(4 rows)
```

Puede obtener el ID de la tabla de la sección DETAIL del mensaje de error para las violaciones de atributos serializables (error 1023).

## STL_UNDONE

Muestra información acerca de las transacciones que se deshicieron.

STL_UNDONE es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_TRANSACTION_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| xact_id | bigint | ID de la transacción deshacer. |
| xact_id_undone | bigint | ID de la transacción que se deshizo. |
| undo_start_ts | marca de tiempo | Hora de inicio de la transacción deshacer. |
| undo_end_ts | marca de tiempo | Hora de finalización de la transacción deshacer. |
| table_id | bigint | ID de la tabla que se vio afectada por la transacción deshacer. |

Consulta de ejemplo

Para ver un registro preciso de todas las transacciones deshacer, escriba el siguiente comando:

```
select xact_id, xact_id_undone, table_id from stl_undone;
```

Este comando devuelve la siguiente salida de ejemplo:

```
 xact_id | xact_id_undone | table_id
---------+----------------+----------
1344 |           1344 |   100192
1326 |           1326 |   100192
1551 |           1551 |   100192
(3 rows)
```

## STL_UNIQUE

Analiza los pasos de ejecución que tienen lugar cuando se utiliza una función DISTINCT en la lista SELECT o cuando se eliminan los duplicados en una consulta UNION o INTERSECT.

STL_UNIQUE es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_UNIQUE solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| slice | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |
| segment | entero | Número que identifica el segmento de la consulta. |
| paso | entero | Paso de la consulta que se ejecutó. |
| starttime | marca de tiempo | Hora en UTC en que comenzó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| endtime | marca de tiempo | Hora en UTC en la que finalizó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| tasknum | entero | Es el número del proceso de tarea de la consulta que se asignó para ejecutar el paso. |
| rows | bigint | Es la cantidad total de filas que se procesaron. |
| `type` | character (6) | El tipo de paso. Los valores válidos son:<br>• HASHED. Indica que el paso utilizó una agregación desordenada agrupada.<br>• PLAIN. Indica que el paso utilizó una agregación escalar no agrupada.<br>• SORTED. Indica que el paso utilizó una agregación ordenada agrupada. |
| is_diskbased | character (1) | Si es true (t) (verdadero), significa que la consulta se realizó como una operación basada en el disco. Si es false (f) (falso), implica que la consulta se realizó en la memoria. |
| slots | entero | Cantidad total de buckets hash. |
| workmem | bigint | Cantidad total de bytes en la memoria funcional que se asignaron a este paso. |
| max_buffers_used | bigint | Cantidad máxima de búferes utilizados en la tabla hash antes de pasar al disco. |
| resizes | entero | Esta información es solo para uso interno. |
| occupied | entero | Esta información es solo para uso interno. |
| flushable | entero | Esta información es solo para uso interno. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| used_unique_prefetching | character (1) | Esta información es solo para uso interno. |
| bytes | biginit | Número de bytes de todas las filas de resultado del paso. |

Consultas de ejemplo

Suponga que ejecuta la siguiente consulta:

```
select distinct eventname
from event order by 1;
```

Si suponemos que el ID de la consulta anterior es 6313, en el siguiente ejemplo se muestra la cantidad de filas que produjo el paso único para cada sector de los segmentos 0 y 1.

```
select query, slice, segment, step, datediff(msec, starttime, endtime) as msec,
 tasknum, rows
from stl_unique where query = 6313
order by query desc, slice, segment, step;
```

```
 query | slice | segment | step | msec | tasknum | rows
-------+-------+---------+------+------+---------+------
  6313 |     0 |       0 |    2 |    0 |      22 |  550
  6313 |     0 |       1 |    1 |  256 |      20 |  145
  6313 |     1 |       0 |    2 |    1 |      23 |  540
  6313 |     1 |       1 |    1 |   42 |      21 |  127
  6313 |     2 |       0 |    2 |    1 |      22 |  540
  6313 |     2 |       1 |    1 |  255 |      20 |  158
  6313 |     3 |       0 |    2 |    1 |      23 |  542
  6313 |     3 |       1 |    1 |   38 |      21 |  146
(8 rows)
```

## STL_UNLOAD_LOG

Registra los detalles de una operación de descarga.

STL_UNLOAD_LOG registra una fila por cada archivo creado por una instrucción UNLOAD. Por ejemplo, si una instrucción UNLOAD crea 12 archivos, STL_UNLOAD_LOG tendrá 12 filas correspondientes.

STL_UNLOAD_LOG es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_UNLOAD_LOG solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_UNLOAD_HISTORY y SYS_UNLOAD_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | El ID de la consulta. |
| slice | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |
| pid | entero | ID de proceso asociado a la instrucción de la consulta. |
| path | character (1280) | La ruta competa del objeto de Amazon S3 para el archivo. |
| start_time | marca de tiempo | Hora de inicio de la transacción. |
| end_time | marca de tiempo | Hora de finalización de la transacción. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| line_count | bigint | Cantidad de líneas (filas) descargadas en el archivo. |
| transfer_size | bigint | Cantidad de bytes transferidos. |
| file_format | character(10) | Formato de archivo no cargado. |

Consulta de ejemplo

Para obtener una lista de los archivos que se escribieron en Amazon S3 mediante un comando UNLOAD, puede llamar a una operación de lista de Amazon S3 una vez que se complete el comando UNLOAD. También puede consultar STL_UNLOAD_LOG.

La siguiente consulta devuelve el nombre de la ruta de los archivos que se crearon a través de un comando UNLOAD para la última consulta completada:

```
select query, substring(path,0,40) as path
from stl_unload_log
where query = pg_last_query_id()
order by path;
```

Este comando devuelve la siguiente salida de ejemplo:

```
 query |             path
-------+---------------------------------------
  2320 | s3://amzn-s3-demo-bucket/venue0000_part_00
  2320 | s3://amzn-s3-demo-bucket/venue0001_part_00
  2320 | s3://amzn-s3-demo-bucket/venue0002_part_00
  2320 | s3://amzn-s3-demo-bucket/venue0003_part_00
(4 rows)
```

## STL_USAGE_CONTROL

La vista STL_USAGE_CONTROL contiene información que se registra cuando se alcanza un límite de uso. Para obtener más información acerca de los límites de uso, consulte Administración de los límites de uso en la Guía de administración de Amazon Redshift.

Solo los superusuarios pueden ver STL_USAGE_CONTROL. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| eventtime | marca de tiempo | Hora (UTC) en que la consulta superó un límite de uso. |
| consulta | entero | Identificador de la consulta. Puede usar este ID para combinar distintas vistas y tablas del sistema. |
| xid | bigint | Identificador de la transacción. |
| pid | entero | Identificador del proceso asociado a la consulta. |
| usage_limit_id | character(40) | Identificador único universal (UUID) generado por Amazon Redshift; por ejemplo, `25d9297e-3e7b-41c8-9f4d-c4b6eb731c09` . |
| feature_type | character (30) | Característica cuyo límite de uso se ha superado. Entre los valores posibles se incluyen: `CONCURRENCY_SCALING` y `SPECTRUM`. |

Consulta de ejemplo

En el siguiente ejemplo de SQL, se devuelve parte de la información que se registra cuando se alcanza un límite de uso.

```
select query, pid, eventtime, feature_type
from stl_usage_control
order by eventtime desc
limit 5;
```

## STL_USERLOG

Registra detalles de los siguientes cambios realizados en un usuario de una base de datos:

- Creación de usuario

- Eliminación de usuario
- Modificación de usuario (cambio de nombre)
- Modificación de usuario (modificación de propiedades)

STL_USERLOG solo es visible para los superusuarios. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_USERLOG. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario afectado por el cambio. |
| nombre de usuario | character (50) | Nombre del usuario afectado por el cambio. |
| olduserna me | character (50) | En el caso de una acción de cambio de nombre, el nombre de usuario original. Para el resto de las acciones, este campo está vacío. |
| acción | character (10) | Acción que ocurrió. Valores válidos:<br>• Modificación<br>• Creación<br>• Drop (Eliminar)<br>• Cambio de nombre |
| usecreate db | entero | Si es true (1), indica que el usuario ha creado privilegios de base de datos. |
| usesuper | entero | Si es true (1), indica que el usuario es un superusuario. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| usecatupd | entero | Si es true (1), indica que el usuario puede actualizar los catálogos de sistema. |
| valuntil | marca de tiempo | Fecha de vencimiento de la contraseña. |
| pid | entero | ID de proceso. |
| xid | bigint | ID de transacción. |
| recordtime | marca de tiempo | Hora en UTC en que comenzó la consulta. |

Consultas de ejemplo

En el siguiente ejemplo, se realizan cuatro acciones del usuario y, luego, se consulta la vista STL_USERLOG.

```
create user userlog1 password 'Userlog1';
alter user userlog1 createdb createuser;
alter user userlog1 rename  to userlog2;
drop user userlog2;

select userid, username, oldusername, action, usecreatedb, usesuper from stl_userlog
 order by recordtime desc;
```

```
 userid |  username | oldusername |  action | usecreatedb | usesuper
--------+-----------+-------------+---------+-------------+----------
    108 | userlog2  |             | drop    |           1 |   1
    108 | userlog2  | userlog1    | rename  |           1 |   1
    108 | userlog1  |             | alter   |           1 |   1
    108 | userlog1  |             | create  |           0 |   0
 (4 rows)
```

# STL_UTILITYTEXT

Captura el texto de los comandos ejecutados en la base de datos que no son SELECT SQL.

Consulte la vista STL_UTILITYTEXT para capturar el siguiente subconjunto de instrucciones SQL que se ejecutaron en el sistema:

- ABORT, BEGIN, COMMIT, END, ROLLBACK
- ANALYZE
- CALL
- CANCEL
- COMMENT
- CREATE, ALTER, DROP DATABASE
- CREATE, ALTER, DROP USER
- EXPLAIN
- GRANT, REVOKE
- LOCK
- RESET
- SET
- SHOW
- TRUNCATE

Consulte también STL_DDLTEXT, STL_QUERYTEXT y SVL_STATEMENTTEXT.

Utilice las columnas STARTTIME y ENDTIME para averiguar qué instrucciones se registraron durante un periodo determinado. Los bloques grandes de texto de SQL se dividen en líneas de 200 caracteres. La columna SEQUENCE identifica los segmentos de texto que pertenecen a una única instrucción.

STL_UTILITYTEXT es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_QUERY_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| xid | bigint | ID de transacción. |
| pid | entero | ID de proceso asociado a la instrucción de la consulta. |
| etiqueta | character(320) | Ya sea el nombre del archivo utilizado para ejecutar la consulta o una etiqueta definida con un comando SET QUERY_GROUP. Si la consulta no se basa en archivos o si no se establece el parámetro QUERY_GROUP, el valor del campo queda en blanco. |
| starttime | marca de tiempo | Hora en UTC en que comenzó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| endtime | marca de tiempo | Hora en UTC en la que finalizó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| sequence | entero | Cuando una única instrucción tiene más de 200 caractere s, se registran filas adicionales para esa instrucción. Secuencia 0 es la primera fila, 1 es la segunda y, así, sucesivamente. |
| texto | character(200) | Texto SQL, en incrementos de 200 caracteres. El campo puede contener caracteres especiales como una barra inversa(\\) y una nueva línea (\n). |

### Consultas de ejemplo

La siguiente consulta devuelve el texto de los comandos "utility" que se ejecutaron el 26 de enero de 2012. En este caso, se ejecutaron algunos comandos SET y un comando SHOW ALL:

```
select starttime, sequence, rtrim(text)
from stl_utilitytext
where starttime like '2012-01-26%'
order by starttime, sequence;

starttime          | sequence |              rtrim
---------------------------+-----+---------------------------------
2012-01-26 13:05:52.529235 |   0 | show all;
2012-01-26 13:20:31.660255 |   0 | SET query_group to ''
2012-01-26 13:20:54.956131 |   0 | SET query_group to 'soldunsold.sql'
...
```

### Reconstruir SQL almacenados

Para reconstruir el SQL almacenado en la columna `text` de STL_UTILITYTEXT, ejecute una instrucción SELECT para crear SQL de 1 o más partes en la columna `text`. Antes de ejecutar el SQL reconstruido, reemplace cualquier caracter especial (\n) con una nueva línea. El resultado de la siguiente instrucción SELECT es filas de SQL reconstruido en el campo `query_statement`.

```
SELECT LISTAGG(CASE WHEN LEN(RTRIM(text)) = 0 THEN text ELSE RTRIM(text) END) WITHIN
 GROUP (ORDER BY sequence) as query_statement
FROM stl_utilitytext GROUP BY xid order by xid;
```

Por ejemplo, la siguiente búsqueda establece el query_group a una cadena de ceros. La búsqueda en si misma es más larga de 200 caracteres y se guarda en varias partes en STL_UTILITYTEXT.

```
set query_group to
 '0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
            000000';
```

En este ejemplo, la búsqueda se guarda en 2 partes (filas) en la columna `text` de STL_UTILITYTEXT.

```
select query, sequence, text
```

```
from stl_utilitytext where query=pg_last_query_id() order by query desc, sequence limit
 10;
```

```
         starttime          | sequence |
                                                 text

----------------------------+----------
+-----------------------------------------------------------------------------------------------
 2019-07-23 22:55:34.926198 |        0 | set query_group to
 '0000000000000000000000000000000000000000000000000000000000000000000000000000000000\n00000000000
 2019-07-23 22:55:34.926198 |        1 |                    000000';
```

Para reconstruir el SQL almacenado en STL_UTILITYTEXT, ejecute el siguiente SQL.

```
select LISTAGG(CASE WHEN LEN(RTRIM(text)) = 0 THEN text ELSE RTRIM(text) END, '')
 within group (order by sequence) AS query_statement
from stl_utilitytext where query=pg_last_query_id();
```

Para usar el SQL reconstruido resultante en su cliente, reemplace cualquier (\n) caracter especial con una nueva línea.

```
                                                query_statement


----------------------------------------------------------------------------------------------
 set query_group to
 '0000000000000000000000000000000000000000000000000000000000000000000000000000000000\n00000000000
                  000000';
```

## STL_VACUUM

Muestra las estadísticas de filas y de bloques de las tablas limpiadas.

En la vista, se muestra información específica respecto a cuándo comenzó y finalizó la operación de limpieza y se especifican los beneficios de ejecutar la operación. Para obtener información acerca de los requisitos para ejecutar este comando, consulte la descripción del comando VACUUM.

STL_VACUUM solo es visible para los superusuarios. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_VACUUM_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| xid | bigint | ID de la transacción para la instrucción VACUUM. Puede unir esta tabla a la vista STL_QUERY para ver las instrucciones SQL individuales que se ejecutan para una transacción VACUUM determinada. Si limpia toda la base de datos, cada tabla se limpia en una transacción independiente. |
| table_id | entero | El ID de la tabla. |
| status | character (30) | Estado de la operación VACUUM para cada tabla. Los valores posibles son los siguientes:<br>• **Started**<br>• **Started Delete Only**<br>• **Started Delete Only (Sorted >= nn%)**<br>Solo se inició la fase eliminar para la operación VACUUM FULL. Se omitió la fase de ordenación porque la tabla ya estaba ordenada en el umbral de ordenación o por encima de este.<br>• **Started Sort Only**<br>• **Started Ranged Partition**<br>• **Started Reindex** |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | • **Finished**<br><br>Hora en la que se completó la operación para la tabla. Para averiguar cuánto tardo la operación de limpieza en una tabla específica, reste la hora de finalización de la hora de inicio de un ID de una transacción y de un ID de tabla en particular.<br>• **Skipped**<br><br>Se omitió la tabla porque ya estaba ordenada por completo y no había filas marcadas para ser eliminadas.<br>• **Skipped (delete only)**<br><br>Se omitió la tabla porque se especificó la opción DELETE ONLY y porque no había filas marcadas para ser eliminadas.<br>• **Skipped (sort only)**<br><br>Se omitió la tabla porque se especificó la opción SORT ONLY y la tabla ya estaba ordenada por completo.<br>• **Skipped (sort only, sorted>=xx%)**<br><br>Se omitió la tabla porque se especificó la opción SORT ONLY y la tabla ya estaba ordenada en el umbral de ordenación o por encima de este.<br>• **Skipped (0 rows)**<br><br>Se omitió la tabla porque estaba vacía.<br>• **VacuumBG**<br><br>Se realizó una operación de limpieza automática en segundo plano. Este estado se antepone a otros estados cuando se realizan automáticamente. Por ejemplo, una operación de limpieza DELETE ONLY |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | realizada automáticamente tendría una fila inicial con el estado `[VacuumBG] Started Delete Only .`<br><br>Para obtener más información acerca de la configuración del umbral de ordenación para el comando VACUUM, consulte VACUUM. |
| rows | bigint | La cantidad real de filas de la tabla más las filas eliminadas que siguen almacenadas en el disco (a la espera de ser limpiadas). En esta columna, se muestra el recuento antes de que se inicie la limpieza de las filas con el estado **`Started`** y el recuento después de la limpieza de las filas con estado **`Finished`**. |
| sortedrows | entero | La cantidad de filas que se ordenan en la tabla. En esta columna, se muestra el recuento antes de que se inicie la limpieza de las filas con el estado **`Started`** en la columna Status (Estado) y el recuento después de la limpieza de las filas con el estado **`Finished`** en esta misma columna. |
| blocks | entero | Cantidad total de bloques de datos que se utilizaron para almacenar datos de tablas antes de la operación de limpieza (filas con el estado **`Started`**) y después de operación de limpieza (columna **`Finished`**). Cada bloque de datos utiliza 1 MB. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| max_merge_partitions | entero | Esta columna se usa para el análisis de rendimiento y representa la cantidad máxima de particiones que la operación de limpieza pueda procesar para la tabla por iteración de fase de fusión. (La limpieza ordena la región desordenada en una o más particiones ordenadas . En función de la cantidad de columnas en la tabla y la configuración actual de Amazon Redshift, la fase de fusión puede procesar una cantidad máxima de particiones en una única iteración de fusión. La fase de fusión funcionará incluso si la cantidad de particiones ordenadas supera la cantidad máxima de particiones de fusión, pero se requerirán más iteraciones de fusión). |
| eventtime | marca de tiempo | Hora en la que comenzó o finalizó la operación de limpieza. |
| reclaimable_rows | bigint | El número de filas recuperables del cutoff_xid actual. Esta columna muestra el número estimado de filas recuperables de Redshift antes de que se iniciara la operación de limpieza para las filas con un estado **`Started`** y el número real de filas recuperables que quedan después de la operación de limpieza para las filas con un estado **`Finished`**. |
| reclaimable_space_mb | bigint | Espacio recuperable en MB para el cutoff_xid actual. Esta columna muestra la cantidad estimada de espacio recuperable de Redshift antes de que se iniciara la operación de limpieza para las filas con un estado **`Started`** y la cantidad real de filas recuperables que quedan después de la operación de limpieza para las filas con un estado **`Finished`**. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| cutoff_xid | bigint | El ID de la transacción de límite para la operación VACUUM. Las transacciones posteriores al límite no se incluyen en la operación VACUUM. |
| is_recluster | entero | Si es 1 (verdadero), la operación VACUUM ejecutó el algoritmo de recluster; si es 0 (falso), no lo hizo. |

Consultas de ejemplo

La siguiente consulta proporciona las estadísticas de limpieza de la tabla 108313. La tabla se limpió por medio de una serie de inserciones y eliminaciones.

```
select xid, table_id, status, rows, sortedrows, blocks, eventtime,
                reclaimable_rows, reclaimable_space_mb
from stl_vacuum where table_id=108313 order by eventtime;

 xid   | table_id | status                  | rows | sortedrows | blocks |  eventtime
         | reclaimable_rows | reclaimable_space_mb
-------+----------+-------------------------+------+------------+--------
+---------------------+-----------------+----------------------
 14294 |   108313 | Started                 | 1950 |        408 |     28 |  2016-05-19
 17:36:01 |              984 |                   17
 14294 |   108313 | Finished                |  966 |        966 |     11 |  2016-05-19
 18:26:13 |                0 |                    0
 15126 |   108313 | Skipped(sorted>=95%)    |  966 |        966 |     11 |  2016-05-19
 18:26:38 |                0 |                    0
```

Al inicio de la operación VACUUM, la tabla contenía 1950 filas almacenadas en 28 bloques de 1 MB. Amazon Redshift estimó que podría recuperar 984 o 17 bloques de espacio en disco, con una operación de limpieza.

En la fila del estado Finished (Finalizado), la columna ROWS muestra un valor de 966 y el valor de la columna BLOCKS es 11, en vez de 28. La operación de limpieza recuperó la cantidad estimada de espacio en disco y no quedó espacio ni filas recuperables una vez finalizada la operación de limpieza.

En la fase de ordenación (transacción 15126), la limpieza pudo omitir la tabla porque las filas se insertaron en el orden de la clave de ordenación.

En el siguiente ejemplo, se muestran las estadísticas de la limpieza SORT ONLY en la tabla SALES (tabla 110116 en este ejemplo) después de una gran operación INSERT:

```
vacuum sort only sales;

select xid, table_id, status, rows, sortedrows, blocks, eventtime
from stl_vacuum order by xid, table_id, eventtime;

xid |table_id|     status      | rows  |sortedrows|blocks|     eventtime
----+--------+-----------------+-------+----------+------+--------------------
...
2925| 110116 |Started Sort Only|1379648|   172456 |  132 | 2011-02-24 16:25:21...
2925| 110116 |Finished         |1379648|  1379648 |  132 | 2011-02-24 16:26:28...
```

## STL_WINDOW

Analiza los pasos de la consulta que realizan las funciones de ventana.

STL_WINDOW es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STL_WINDOW solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| slice | entero | Es el número que identifica el sector donde se estaba ejecutando la consulta. |
| segment | entero | Número que identifica el segmento de la consulta. |
| paso | entero | Paso de la consulta que se ejecutó. |
| starttime | marca de tiempo | Hora en UTC en que comenzó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| endtime | marca de tiempo | Hora en UTC en la que finalizó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |
| tasknum | entero | Es el número del proceso de tarea de la consulta que se asignó para ejecutar el paso. |
| rows | bigint | Es la cantidad total de filas que se procesaron. |
| is_diskbased | character (1) | Si es true (t) (verdadero), significa que la consulta se realizó como una operación basada en el disco. Si es false (f) (falso), implica que la consulta se realizó en la memoria. |
| workmem | bigint | Cantidad total de bytes en la memoria funcional que se asignaron a este paso. |

Consultas de ejemplo

El siguiente ejemplo devuelve los resultados de la función de ventana para el sector 0 y el segmento 3.

```
select query, tasknum, rows, is_diskbased, workmem
from stl_window
where slice=0 and segment=3;
```

```
 query | tasknum | rows | is_diskbased | workmem
-------+---------+------+--------------+----------
 86326 |      36 | 1857 | f            | 95256616
   705 |      15 | 1857 | f            | 95256616
 86399 |      27 | 1857 | f            | 95256616
   649 |      10 |    0 | f            | 95256616
(4 rows)
```

## STL_WLM_ERROR

Registra todos los errores relacionados con la Workload Management (WLM, Administración de cargas de trabajo) a medida que se producen.

STL_WLM_ERROR es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| recordtime | marca de tiempo | Hora en la que ocurrió el error. |
| pid | entero | ID de proceso que generó el error. |
| error_string | character (256) | Descripción del error. |

# STL_WLM_RULE_ACTION

Registra detalles de las acciones que surgen de las reglas de supervisión de consultas de la Workload Management (WLM, Administración de cargas de trabajo) asociadas a colas definidas por el usuario. Para obtener más información, consulte Reglas de monitoreo de consultas de WLM.

STL_WLM_RULE_ACTION es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | Usuario que ejecutó la consulta. |
| consulta | entero | ID de la consulta. |
| service_class | entero | ID para la clase de servicio. Las colas de consultas se definen en la configuración de WLM. Las clases de servicios mayores que 5 son colas definidas por el usuario. |
| rule | character (256) | Nombre de una regla de supervisión de consultas. |
| acción | character (256) | Acción resultante. Los valores posibles son los siguientes:<br>• log<br>• hop(reassign)<br>• hop(restart)<br>• abort<br>• change_query_priority<br>• none<br><br>El valor none indica que se cumplieron los predicados de la regla, pero que la acción se reemplazó por otra regla con una acción de mayor prioridad. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| recordtime | marca de tiempo | Hora en la que se registró la acción en UTC. |
| action_value | character (256) | Si `action` es `change_query_priority` , los valores posibles son `highest`, `high`, `normal`, `low` y `lowest`.<br><br>Si `action` es `log`, hop o `abort`, el valor está vacío. |
| service_class_name | character(64) | El nombre de la clase de servicio. |

Consultas de ejemplo

En el siguiente ejemplo, se encuentran consultas que fueron detenidas por una regla de supervisión de consultas.

```
Select query, rule
from stl_wlm_rule_action
where action = 'abort'
order by query;
```

## STL_WLM_QUERY

Presenta un registro de cada intento de ejecución de una consulta en una clase de servicio gestionada por la Workload Management (WLM, Administración de cargas de trabajo).

STL_WLM_QUERY es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_QUERY_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| xid | entero | ID de la transacción de la consulta o de la subconsulta. |
| tarea | entero | ID utilizado para hacer un seguimiento de una consulta a través del administrador de cargas de trabajo. Puede estar asociado a distintos ID de consultas. Si se reinicia una consulta, esta se asigna a un ID de consulta nuevo, pero no a un ID de tarea nuevo. |
| consulta | entero | ID de la consulta. Si se reinicia una consulta, esta se asigna a un ID de consulta nuevo, pero no a un ID de tarea nuevo. |
| service_class | entero | ID para la clase de servicio. Para obtener una lista de ID de clase de servicio, consulte ID de la clase de servicio WLM. |
| slot_count | entero | Número de ranuras de consulta de WLM que utiliza una consulta según el nivel de simultaneidad establecido para la cola. El valor predeterminado es 1. Para obtener más información, consulte wlm_query_slot_count. |
| service_class_start_time | marca de tiempo | Hora en la que se asignó la consulta a la clase de servicio. Esta hora está en la zona horaria UTC. |
| queue_start_time | marca de tiempo | Hora en la que la consulta entró en la cola para la clase de servicio. Esta hora está en la zona horaria UTC. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| queue_end_time | marca de tiempo | Hora en la que la consulta salió de la cola para la clase de servicio. Esta hora está en la zona horaria UTC. |
| total_queue_time | bigint | Cantidad total de microsegundos que estuvo la consulta en la cola |
| exec_start_time | marca de tiempo | Hora en la que se comenzó a ejecutar la consulta en la clase de servicio. Esta hora está en la zona horaria UTC. |
| exec_end_time | marca de tiempo | Hora en la que se terminó de ejecutar la consulta en la clase de servicio. Esta hora está en la zona horaria UTC. |
| total_exec_time | bigint | Cantidad de microsegundos que la consulta estuvo en ejecución. |
| service_class_end_time | marca de tiempo | Hora en la que la consulta salió de la clase de servicio. Esta hora está en la zona horaria UTC. |
| final_state | character(16) | Reservado para el uso del sistema. |
| est_peak_mem | bigint | Reservado para el uso del sistema. |
| query_priority | char(20) | La prioridad de la consulta. Los valores posibles son `n/a`, `lowest`, `low`, `normal`, `high` y `highest`, donde `n/a` significa que no se admite la prioridad de la consulta. |
| service_class_name | character(64) | Es el nombre de la clase de servicio. Para obtener más información sobre las clases de servicio, consulte las tablas y vistas de sistema de WLM. |

Consultas de ejemplo

Obtención del tiempo promedio de las consultas de las colas y en ejecución

Las siguientes consultas muestran la configuración actual de las clases de servicio mayores que 4. Para obtener una lista de ID de clase de servicio, consulte ID de la clase de servicio WLM.

La siguiente consulta devuelve el tiempo promedio (en microsegundos) que pasa cada consulta en las colas de consultas y el tiempo que están en ejecución para cada clase de servicio.

```
select service_class as svc_class, count(*),
avg(datediff(microseconds, queue_start_time, queue_end_time)) as avg_queue_time,
avg(datediff(microseconds, exec_start_time, exec_end_time )) as avg_exec_time
from stl_wlm_query
where service_class > 4
group by service_class
order by service_class;
```

Esta consulta devuelve el siguiente ejemplo de salida:

```
 svc_class | count | avg_queue_time | avg_exec_time
-----------+-------+----------------+---------------
         5 | 20103 |              0 |         80415
         5 |  3421 |          34015 |        234015
         6 |    42 |              0 |        944266
         7 |   196 |           6439 |       1364399
(4 rows)
```

Obtención del tiempo máximo de las consultas de las colas y en ejecución

La siguiente consulta devuelve el tiempo máximo (en microsegundos) que pasa una consulta en alguna cola de consultas y el tiempo que está en ejecución para cada clase de servicio.

```
select service_class as svc_class, count(*),
max(datediff(microseconds, queue_start_time, queue_end_time)) as max_queue_time,
max(datediff(microseconds, exec_start_time, exec_end_time )) as max_exec_time
from stl_wlm_query
where svc_class > 5
group by service_class
order by service_class;
```

```
 svc_class | count | max_queue_time | max_exec_time
```

```
-----------+-------+----------------+---------------
         6 |    42 |              0 |       3775896
         7 |   197 |          37947 |      16379473
(4 rows)
```

## Tablas STV para datos de instantáneas

Las tablas de STV son tablas de sistema virtuales que tienen instantáneas de los datos actuales del sistema.

Temas

- STV_ACTIVE_CURSORS
- STV_BLOCKLIST
- STV_CURSOR_CONFIGURATION
- STV_DB_ISOLATION_LEVEL
- STV_EXEC_STATE
- STV_INFLIGHT
- STV_LOAD_STATE
- STV_LOCKS
- STV_ML_MODEL_INFO
- STV_MV_DEPS
- STV_MV_INFO
- STV_NODE_STORAGE_CAPACITY
- STV_PARTITIONS
- STV_QUERY_METRICS
- STV_RECENTS
- STV_SESSIONS
- STV_SLICES
- STV_STARTUP_RECOVERY_STATE
- STV_TBL_PERM
- STV_TBL_TRANS
- STV_WLM_CLASSIFICATION_CONFIG
- STV_WLM_QMR_CONFIG

- STV_WLM_QUERY_QUEUE_STATE
- STV_WLM_QUERY_STATE
- STV_WLM_QUERY_TASK_STATE
- STV_WLM_SERVICE_CLASS_CONFIG
- STV_WLM_SERVICE_CLASS_STATE
- STV_XRESTORE_ALTER_QUEUE_STATE

## STV_ACTIVE_CURSORS

STV_ACTIVE_CURSORS muestra los detalles de los cursores que están abiertos actualmente. Para obtener más información, consulte DECLARE.

Todos los usuarios pueden ver STV_ACTIVE_CURSORS. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema. Un usuario solo puede ver los cursores que él mismo haya abierto. Un superusuario puede ver todos los cursores.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| nombre | characte (256) | Nombre del cursor. |
| xid | bigint | Contexto de la transacción. |
| pid | entero | Proceso principal que ejecuta la consulta. |
| starttime | marca de tiempo | El momento en el que se declaró el cursor. |
| row_count | bigint | Cantidad de filas en el conjunto de resultados del cursor. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| byte_count | bigint | Cantidad de bytes en el conjunto de resultados del cursor. |
| fetched_rows | bigint | Cantidad de filas que se obtienen actualmente del conjunto de resultados del cursor. |

## STV_BLOCKLIST

STV_BLOCKLIST tiene la cantidad de bloques de 1 MB de disco que utiliza cada sector, tabla o columna en una base de datos.

Utilice consultas de agregación con STV_BLOCKLIST, como se muestra en los siguientes ejemplos, para determinar la cantidad de bloques de 1 MB de disco asignados para cada base de datos, tabla, sector o columna. También puede utilizar STV_PARTITIONS para obtener información resumida acerca de la utilización del disco.

Solo los superusuarios pueden ver STV_BLOCKLIST. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> STV_BLOCKLIST solo registra los bloques que pertenecen al clúster aprovisionado o al espacio de nombres sin servidor. Si la base de datos incluye bloques compartidos de un productor de recursos compartidos de datos, esos bloques no se incluyen en STV_BLOCKLIST. Para obtener más información acerca de los recursos compartidos de datos, vaya a Uso compartido de datos en Amazon Redshift.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| slice | entero | Sector del nodo. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| col | entero | Índice con base cero para la columna. Cada tabla que cree tiene tres columnas ocultas anexadas: INSERT_XID, DELETE_XID y ROW_ID (OID). Una tabla con 3 columnas definidas por el usuario tiene 6 columnas reales, y las columnas definidas por el usuario se enumeran internamente como 0, 1 y 2. Las columnas INSERT_XID, DELETE_XID y ROW_ID se enumeran 3, 4 y 5, respectivamente, en este ejemplo. |
| tbl | entero | ID de la tabla para la tabla de la base de datos. |
| blocknum | entero | ID para el bloque de datos. |
| num_values | entero | Cantidad de valores contenidos en el bloque. |
| extended_limits | entero | Para uso interno. |
| minvalue | bigint | Valor mínimo de datos del bloque. Almacena los primeros ocho caracteres como un entero de 64 bits para datos no numéricos. Se utiliza para explorar el disco. |
| maxvalue | bigint | Valor máximo de dato del bloque. Almacena los primeros ocho caractere s como un entero de 64 bits para datos no numéricos. Se utiliza para explorar el disco. |
| sb_pos | entero | Identificador interno de Amazon Redshift para la posición del superbloq ue en el disco. |
| pinned | entero | Indica si el bloque se conectó o no a la memoria como parte de la carga previa: 0 = false; 1 = true. La opción predeterminada es falso. |
| on_disk | entero | Indica si el bloque se almacenó automáticamente o no en el disco: 0 = false; 1 = true. La opción predeterminada es falso. |
| modified | entero | Indica si el bloque se modificó o no: 0 = false; 1 = true. La opción predeterminada es falso. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| hdr_modified | entero | Indica si el encabezado del bloque se modificó o no: 0 = false; 1 = true. La opción predeterminada es falso. |
| unsorted | entero | Indica si el bloque está desordenado o no: 0 = false; 1 = true. El valor predeterminado es verdadero. |
| tombstone | entero | Para uso interno. |
| preferred_diskno | entero | Cantidad de discos en que debe estar el bloque, excepto que el disco tenga una falla. Una vez arreglado el disco, el bloque volverá a ese disco. |
| temporary | entero | Indica si el bloque tiene o no datos temporales, como una tabla temporal o resultados intermedios de consulta: 0 = false; 1 = true. La opción predeterminada es falso. |
| newblock | entero | Indica si un bloque es o no nuevo (true) o si nunca se guardó en el disco (false): 0 = false; 1 = true. |
| num_readers | entero | Cantidad de referencias en cada bloque. |
| flags | entero | Indicadores internos de Amazon Redshift para el encabezado del bloque. |

Consultas de ejemplo

STV_BLOCKLIST tiene una fila por cada bloque del disco designado, por lo que una consulta que selecciona todas las filas posiblemente devuelva una gran cantidad de filas. Le recomendamos usar solo las consultas de agregación con STV_BLOCKLIST.

La vista SVV_DISKUSAGE proporciona información similar en un formato más sencillo de usar. Sin embargo, en el siguiente ejemplo se demuestra un uso de la tabla STV_BLOCKLIST.

Para determinar la cantidad de bloques de 1 MB que utiliza cada columna de la tabla VENUE, escriba la siguiente consulta:

```
select col, count(*)
from stv_blocklist, stv_tbl_perm
where stv_blocklist.tbl = stv_tbl_perm.id
and stv_blocklist.slice = stv_tbl_perm.slice
and stv_tbl_perm.name = 'venue'
group by col
order by col;
```

Esta consulta devuelve la cantidad de bloques de 1 MB asignados a cada columna de la tabla VENUE, tal como se muestra en los siguientes datos de ejemplo:

```
 col | count
-----+-------
   0 |  4
   1 |  4
   2 |  4
   3 |  4
   4 |  4
   5 |  4
   7 |  4
   8 |  4
(8 rows)
```

En la siguiente consulta, se muestra si una tabla está distribuida en todos los sectores:

```
select trim(name) as table, stv_blocklist.slice, stv_tbl_perm.rows
from stv_blocklist,stv_tbl_perm
where stv_blocklist.tbl=stv_tbl_perm.id
and stv_tbl_perm.slice=stv_blocklist.slice
and stv_blocklist.id > 10000 and name not like '%#m%'
and name not like 'systable%'
group by name, stv_blocklist.slice, stv_tbl_perm.rows
order by 3 desc;
```

Esta consulta produce el siguiente ejemplo de salida, que muestra una distribución uniforme de los datos en la tabla que tiene la mayor cantidad de filas:

```
table   | slice | rows
----------+-------+-------
listing  |    13 | 10527
listing  |    14 | 10526
listing  |     8 | 10526
listing  |     9 | 10526
listing  |     7 | 10525
listing  |     4 | 10525
listing  |    17 | 10525
listing  |    11 | 10525
listing  |     5 | 10525
listing  |    18 | 10525
listing  |    12 | 10525
listing  |     3 | 10525
listing  |    10 | 10525
listing  |     2 | 10524
listing  |    15 | 10524
listing  |    16 | 10524
listing  |     6 | 10524
listing  |    19 | 10524
listing  |     1 | 10523
listing  |     0 | 10521
...
(180 rows)
```

La siguiente consulta determina si hay algún bloque del tipo tombstone confirmado en el disco:

```
select slice, col, tbl, blocknum, newblock
from stv_blocklist
where  tombstone > 0;

slice | col |   tbl  | blocknum | newblock
-------+-----+--------+----------+----------
4     |  0  | 101285 |    0     |   1
4     |  2  | 101285 |    0     |   1
4     |  4  | 101285 |    1     |   1
5     |  2  | 101285 |    0     |   1
5     |  0  | 101285 |    0     |   1
5     |  1  | 101285 |    0     |   1
5     |  4  | 101285 |    1     |   1
...
(24 rows)
```

## STV_CURSOR_CONFIGURATION

STV_CURSOR_CONFIGURATION muestra las restricciones de la configuración del cursor. Para obtener más información, consulte Restricciones del cursor.

Solo los superusuarios pueden ver STV_CURSOR_CONFIGURATION. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| current_cursor_count | entero | Cantidad de cursores que están abiertos actualmente. |
| max_diskspace_usable | entero | Cantidad de espacio disponible en el disco para cursores, en megabytes. Esta restricción se basa en el tamaño máximo del conjunto de resultados de cursores del clúster. |
| current_diskspace_used | entero | Cantidad de espacio del disco que utilizan actualmente los cursores, en megabytes. |

## STV_DB_ISOLATION_LEVEL

STV_DB_ISOLATION_LEVEL muestra el nivel de aislamiento actual de las bases de datos. Para obtener más información acerca de los niveles de aislamiento, consulte CREATE DATABASE.

STV_DB_ISOLATION_LEVEL es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| db_name | characte (128) | El nombre de la base de datos. |
| isolation _level | characte (20) | El nivel de aislamiento de la base de datos. Entre los valores posibles se incluyen: `Serializable` y `Snapshot Isolation` . |

## STV_EXEC_STATE

Utilice la tabla STV_EXEC_STATE para encontrar información acerca de las consultas y los pasos de consultas que se están ejecutando de forma activa en los nodos de computación.

Esta información suele utilizarse solo para solucionar problemas de ingeniería. Las vistas SVV_QUERY_STATE y SVL_QUERY_SUMMARY extraen la información de STV_EXEC_STATE.

Todos los usuarios pueden ver STV_EXEC_STATE. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. Se puede usar para combinar distintas vistas y tablas del sistema. |
| slice | entero | Sección del nodo donde se completa el paso. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| segment | entero | Segmento de la consulta que se ejecutó. El segmento de una consulta es una serie de pasos. |
| paso | entero | Paso del segmento de la consulta que se completó. Un paso es la unidad más pequeña que realiza una consulta. |
| starttime | marca de tiempo | Tiempo en el que se ejecutó el paso. |
| currenttime | marca de tiempo | Hora actual. |
| tasknum | entero | Proceso de tarea de la consulta que se asigna para completar el paso. |
| rows | bigint | Cantidad de filas procesadas. |
| bytes | bigint | Cantidad de bytes procesados. |
| etiqueta | char (256) | Etiqueta del paso, que se compone del nombre de un paso de la consulta y, si corresponde, el ID y el nombre de la tabla (por ejemplo, `scan tbl=100448 name =user`). Los ID de tabla de tres dígitos suelen referirse a exámenes de tablas transitorias. Cuando vea `tbl=0`, por lo general se refiere a un examen de un valor constante. |
| is_diskbased | char(1) | Indique si este paso de la consulta se ejecutó como una operación basada en el disco: true (verdadero) (**`t`**) o false (falso) (**`f`**). Solo determinados pasos, como hash, ordenar y agregar pasos, pueden ir al disco. Muchos tipos de pasos se completan siempre en la memoria. |
| workmem | bigint | Cantidad de bytes de memoria funcional asignada a este paso. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| num_parts | entero | Cantidad de particiones en que se divide una tabla hash durante un paso hash. Un número positivo en esta columna no significa que el paso hash se haya ejecutado como una operación basada en el disco. Controle el valor en la columna IS_DISKBASED para saber si el paso hash se basó en el disco. |
| is_rrscan | char (1) | Si es true (**t**), indica que se utilizó un examen de rango restringido en el paso. El valor predeterminado es false (**f**). |
| is_delayed_scan | char(1) | Si es true (**t**), indica que en el paso se utilizó un examen diferido. El valor predeterminado es false (**f**). |

Consultas de ejemplo

En lugar de consultar la tabla STV_EXEC_STATE directamente, Amazon Redshift recomienda consultar SVL_QUERY_SUMMARY o SVV_QUERY_STATE para obtener la información de STV_EXEC_STATE en un formato más fácil de usar. Consulte la tabla SVL_QUERY_SUMMARY o la documentación de la tabla SVV_QUERY_STATE para obtener más detalles.

## STV_INFLIGHT

Utilice la tabla STV_INFLIGHT para determinar cuáles son las consultas que se están ejecutando en el clúster. Si está solucionando problemas, resulta útil para comprobar el estado de las consultas de larga duración.

STV_INFLIGHT no muestra las consultas únicas del nodo principal. Para obtener más información, consulte Funciones específicas del nodo principal. Todos los usuarios pueden ver STV_INFLIGHT. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_QUERY_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Solución de problemas con STV_INFLIGHT

Si utiliza STV_INFLIGHT para solucionar problemas de rendimiento de una consulta o de una colección de consultas, tenga en cuenta lo siguiente:

- Las transacciones abiertas de larga duración suelen aumentar la carga. Estas transacciones abiertas pueden dar lugar a tiempos de ejecución más largos para otras consultas.
- Los trabajos COPY y ETL de larga duración pueden afectar a otras consultas que se ejecuten en el clúster, si están consumiendo muchos recursos de computación. En la mayoría de los casos, trasladar estos trabajos de larga duración a momentos de poco uso aumenta el rendimiento de las cargas de trabajo de creación de informes o análisis.
- Existen vistas que proporcionan información relacionada con STV_INFLIGHT. Entre ellas se incluye STL_QUERYTEXT, que captura el texto de consulta de los comandos SQL, y SVV_QUERY_INFLIGHT, que une STV_INFLIGHT a STL_QUERYTEXT. También puede utilizar STV_RECENTS con STV_INFLIGHT para la solución de problemas. Por ejemplo, STV_RECENTS puede indicar si determinadas consultas se encuentran en estado En ejecución o Terminado. La combinación de esta información con los resultados de STV_INFLIGHT puede darle más información sobre las propiedades de una consulta y el impacto en los recursos de computación.

También puede monitorear las consultas en ejecución mediante la consola de Amazon Redshift.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| slice | entero | Sector donde se ejecuta la consulta. |
| consulta | entero | ID de la consulta. Se puede usar para combinar distintas vistas y tablas del sistema. |
| etiqueta | character(320) | Ya sea el nombre del archivo utilizado para ejecutar la consulta o una etiqueta definida con un comando SET QUERY_GROUP. Si la consulta no se basa en archivos o |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | si no se establece el parámetro QUERY_GROUP, el valor del campo queda en blanco. |
| xid | bigint | ID de transacción. |
| pid | entero | ID de proceso. Todas las consultas en una sesión se ejecutan en el mismo proceso, por lo que este valor permanece constante si ejecuta una serie de consultas en la misma sesión. Puede utilizar esta columna para combinar la tabla STL_ERROR. |
| starttime | marca de tiempo | La hora en que comenzó la consulta. |
| texto | character(100) | Texto de la consulta, truncado en 100 caracteres si la instrucción supera ese límite. |
| suspended | entero | Indica si la consulta está suspendida o no: 0 = false; 1 = true. |
| insert_pristine | entero | Indica si las consultas de escritura pueden o pudieron ejecutarse mientras que la consulta actual se ejecuta o ejecutaba: 1 = no se permiten consultas de escritura; 0 = se permiten las consultas de escritura. Esta columna está diseñada para ser utilizada en la depuración. |
| concurrency_scaling_status | entero | Indica si la consulta se ejecutó en un clúster principal o en un clúster de escalado de simultaneidad. Los valores posibles son los siguientes:<br><br>0: se ejecutó en el clúster principal<br><br>1: se ejecutó en un clúster de escalado de simultaneidad |

Consultas de ejemplo

Para ver todas las consultas activas que se están ejecutando actualmente en la base de datos, escriba la siguiente consulta:

```
select * from stv_inflight;
```

En el siguiente ejemplo de salida, se muestra dos consultas que se están ejecutando actualmente, incluida la misma consulta STV_INFLIGHT y una consulta que se ejecutó desde un script denominado `avgwait.sql`:

```
select slice, query, trim(label) querylabel, pid,
starttime, substring(text,1,20) querytext
from stv_inflight;

slice|query|querylabel | pid |        starttime         |      querytext
-----+-----+-----------+-----+--------------------------+--------------------
1011 |  21 |           | 646 |2012-01-26 13:23:15.645503|select slice, query,
1011 |  20 |avgwait.sql| 499 |2012-01-26 13:23:14.159912|select avg(datediff(
(2 rows)
```

La siguiente consulta selecciona varias columnas, como concurrency_scaling_status. Esta columna indica si las consultas se están enviando al clúster de escalado de simultaneidad. Si el valor es 1 para algunos resultados, es una indicación de que se están utilizando recursos de computación con escalado de simultaneidad. Para obtener más información, consulte Escalado de simultaneidad.

```
select userid,
query,
pid,
starttime,
text,
suspended,
concurrency_scaling_status
 from STV_INFLIGHT;
```

La salida de ejemplo muestra el envío de una consulta al clúster de escalado de simultaneidad.

```
 query  | pid     |        starttime           |   text                | suspended
 |  concurrency_scaling_status
--------+---------
+---------------------------|-----------------------|--------------|----------------------
```

```
1234567 | 123456  | 2012-01-26 13:23:15.645503 | select userid, query...  0
    1
2345678 | 234567  | 2012-01-26 13:23:14.159912 | select avg(datediff(...  0
    0
(2 rows)
```

Para obtener más consejos sobre la resolución de problemas de rendimiento de las consultas, consulte Solución de problemas de consultas.

## STV_LOAD_STATE

Utilice la tabla STV_LOAD_STATE para encontrar información acerca del estado actual de las instrucciones COPY en curso.

El comando COPY actualiza esta tabla cada vez que se carga un millón de registros.

Todos los usuarios pueden ver STV_LOAD_STATE. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| sesión | entero | El PID de la sesión del proceso que hace la carga. |
| consulta | entero | ID de la consulta. Se puede usar para combinar distintas vistas y tablas del sistema. |
| slice | entero | Número del sector del nodo. |
| pid | entero | ID de proceso. Todas las consultas en una sesión se ejecutan en el mismo proceso, por lo que este valor permanece constante si ejecuta una serie de consultas en la misma sesión. |
| recordtime | marca de tiempo | Hora en la que se graba el registro. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| bytes_to_load | bigint | Cantidad total de bytes que cargará este sector. Este valor es igual a 0 si los datos que se cargan están comprimidos. |
| bytes_loaded | bigint | Cantidad de bytes cargados por este sector. Si los datos que se cargan están comprimidos, este valor corresponde a la cantidad de datos cargados una vez que se descompri men los datos. |
| bytes_to_load_comp ressed | bigint | Cantidad total de bytes de datos comprimidos que cargará este sector. Este valor es igual a 0 si los datos que se cargan no están comprimidos. |
| bytes_loaded_compr essed | bigint | Cantidad de bytes de datos comprimidos cargados por este sector. Este valor es igual a 0 si los datos que se cargan no están comprimidos. |
| lines | entero | Cantidad de líneas cargadas por este sector. |
| num_files | entero | Cantidad de filas que cargará este sector. |
| num_files_complete | entero | Cantidad de filas cargadas por este sector. |
| current_file | character (256) | Nombre de la fila que está cargando este sector. |
| pct_complete | entero | Porcentaje de la carga de datos completada por este sector. |

Consulta de ejemplo

Para ver el progreso de cada sector de un comando COPY, escriba la siguiente consulta. En este ejemplo, se usa la función PG_LAST_COPY_ID() para recuperar información para el último comando COPY.

```
select slice , bytes_loaded, bytes_to_load , pct_complete from stv_load_state where
 query = pg_last_copy_id();
```

```
 slice | bytes_loaded | bytes_to_load | pct_complete
-------+--------------+---------------+--------------
     2 |            0 |             0 |            0
     3 |     12840898 |      39104640 |           32
(2 rows)
```

## STV_LOCKS

Utilice la tabla STV_LOCKS para ver todas las actualizaciones de las tablas en la base de datos.

Amazon Redshift bloquea tablas para evitar que dos usuarios actualicen la misma tabla al mismo tiempo. Mientras la tabla STV_LOCKS muestra todas las actualizaciones actuales de las tablas, consulte la tabla STL_TR_CONFLICT para ver un registro de los conflictos de bloqueos. Utilice la vista SVV_TRANSACTIONS para identificar las transacciones abiertas y bloquear los problemas de contención.

Solo los superusuarios pueden ver STV_LOCKS. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| table_id | bigint | ID de la tabla para la tabla que adquiere el bloqueo. |
| last_commit | marca de tiempo | La marca temporal de la última confirmación en la tabla. |
| last_update | marca de tiempo | La marca temporal de la última actualización de la tabla. |
| lock_owner | bigint | ID de la transacción asociado al bloqueo. |
| lock_owner_pid | bigint | ID del proceso asociado al bloqueo. |
| lock_owner_start_ts | marca de tiempo | Marca temporal para la hora de inicio de la transacción. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| lock_owner_end_ts | marca de tiempo | Marca temporal para la hora de finalización de la transacción. |
| lock_status | character (22) | Estado del proceso que está esperando o reteniendo un bloqueo. |

Consulta de ejemplo

Para ver todos los bloqueos que se están realizando en las transacciones actuales, escriba el siguiente comando:

```
select table_id, last_update, lock_owner, lock_owner_pid from stv_locks;
```

Esta consulta devuelve el siguiente ejemplo de salida, en el que se muestran tres bloqueos actualmente activos:

```
 table_id |        last_update         | lock_owner | lock_owner_pid
----------+----------------------------+------------+----------------
100004  | 2008-12-23 10:08:48.882319 |       1043 |           5656
100003  | 2008-12-23 10:08:48.779543 |       1043 |           5656
100140  | 2008-12-23 10:08:48.021576 |       1043 |           5656
(3 rows)
```

## STV_ML_MODEL_INFO

Información acerca del estado actual del modelo de machine learning.

Todos los usuarios pueden ver STV_ML_MODEL_INFO. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| schema_name | char(128) | El espacio de nombres del modelo. |
| user_name | char(128) | El propietario del modelo. |
| model_name | char(128) | El nombre del modelo. |
| life_cycle | char(20) | El estado del ciclo de vida del modelo. |
| is_refreshable | entero | Indica si el estado del modelo se puede actualizar en caso de que las tablas y las columnas originales en la consulta de formación sigan existiendo y de que el usuario aún cuente con los permisos necesarios para usarlas. Los valores posibles son 1 (actualizable) y 0 (no actualizable). |
| model_state | char(128) | El estado actual del modelo. |

Consulta de ejemplo

La siguiente consulta muestra el estado actual de los modelos de machine learning.

```
SELECT schema_name, model_name, model_state
FROM stv_ml_model_info;

 schema_name |        model_name            |             model_state
-------------+------------------------------+-------------------------------------
 public      | customer_churn_auto_model    | Train Model On SageMaker In Progress
 public      | customer_churn_xgboost_model | Model is Ready
(2 row)
```

## STV_MV_DEPS

En la tabla STV_MV_DEPS, se muestran las dependencias de vistas materializadas con respecto a otras vistas materializadas de Amazon Redshift.

Para obtener más información acerca de las vistas materializadas, consulte Vistas materializadas en Amazon Redshift.

Todos los usuarios pueden ver STV_MV_DEPS. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver las vistas materializadas que residen en los esquemas a los que tienen acceso. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas del sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| db_name | char(128) | La base de datos que contiene la vista materiali zada que se especificó. |
| esquema | char(128) | Esquema de la vista materializada. |
| nombre | char(128) | El nombre de la vista materializada. |
| ref_schema | char(128) | El esquema del cual depende la vista materiali zada. |
| ref_name | char(128) | El nombre del cual depende la vista materiali zada. |
| ref_datab ase_name | char(128) | Nombre de la base de datos de la que depende esta vista materializada. |

Consulta de ejemplo

La siguiente consulta devuelve una fila de salida que indica que la vista materializada `mv_over_foo` utiliza la vista materializada `mv_foo` en su definición como una dependencia.

```
CREATE SCHEMA test_ivm_setup;
CREATE TABLE test_ivm_setup.foo(a INT);
CREATE MATERIALIZED VIEW test_ivm_setup.mv_foo AS SELECT * FROM test_ivm_setup.foo;
CREATE MATERIALIZED VIEW test_ivm_setup.mv_over_foo AS SELECT * FROM
 test_ivm_setup.mv_foo;
```

```
SELECT * FROM stv_mv_deps;

 db_name | schema          | name        |   ref_schema   | ref_name |
 ref_database_name
---------+-----------------+-------------+----------------+----------
+------------------
 dev     | test_ivm_setup  | mv_over_foo | test_ivm_setup | mv_foo   | dev
```

## STV_MV_INFO

La tabla STV_MV_INFO contiene una fila para cada vista materializada, si los datos están o no obsoletos, e información de estado.

Para obtener más información acerca de las vistas materializadas, consulte Vistas materializadas en Amazon Redshift.

Todos los usuarios pueden ver STV_MV_INFO. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver las vistas materializadas que residen en los esquemas a los que tienen acceso. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas del sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| db_name | char(128) | Base de datos que contiene la vista materializada. |
| esquema | char(128) | El esquema de la base de datos. |
| nombre | char(128) | El nombre de la vista materializada. |
| updated_upto_xid | bigint | Se reserva para uso interno. |
| is_stale | char(1) | Una `t` indica que la vista materializada está obsoleta. Una vista materializada obsoleta es aquella en la que las tablas base se han actualizado pero la vista materializada no. Esta informaci |

| Nombre de la columna | Tipo de datos: | Descripción |
| --- | --- | --- |
| | | ón puede no ser exacta si no se ha ejecutado una actualización desde el último reinicio.<br><br>La columna `is_stale` siempre se establece en `t` si la vista materializada depende de una función mutable. Una función mutable devuelve un resultado diferente cuando se le da el mismo argumento o argumentos. Por ejemplo, la mayoría de las funciones que devuelven una fecha o una marca de tiempo son funciones mutables. |
| owner_user_name | char(128) | El usuario propietario de la vista materializada. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| estado | entero | El estado de la vista materializada es el siguiente:<br>• 0: la vista materializada se vuelve a calcular por completo cada vez que se actualiza.<br>• 1: la vista materializada es progresiva.<br>• 101: la vista materializada no se puede actualizar debido a que se ha eliminado una columna. Esta restricción se aplica incluso si la columna no se utiliza en la vista materializada.<br>• 102: la vista materializada no se puede actualizar debido al cambio en el tipo de una columna. Esta restricción se aplica incluso si la columna no se utiliza en la vista materializada.<br>• 103: la vista materializada no se puede actualizar debido a que se ha cambiado el nombre de una tabla.<br>• 104: la vista materializada no se puede actualizar debido a que se ha cambiado el nombre de una columna. Esta restricción se aplica incluso si la columna no se utiliza en la vista materializada.<br>• 105: la vista materializada no se puede actualizar debido a que se ha cambiado el nombre de un esquema. |
| autorewrite | char(1) | `t` indica que la vista materializada reúne los requisitos necesarios para la reescritura automática de las consultas. |
| autorefresh | char(1) | `t` indica que la vista materializada se puede actualizar de manera automática. |

Consulta de ejemplo

Para ver el estado de todas las vistas materializadas, ejecute la siguiente consulta.

```
select * from stv_mv_info;
```

Esta consulta devuelve el siguiente ejemplo de salida.

```
 db_name |       schema       |   name   | updated_upto_xid | is_stale | owner_user_name
 | state | autorefresh | autorewrite
---------+--------------------+---------+------------------+----------
+-----------------+-------+-------------+------------
 dev     | test_ivm_setup     | mv       |             1031 | f        | catch-22
 |     1 |           1 |           0
 dev     | test_ivm_setup     | old_mv   |              988 | t        | lotr
 |     1 |           0 |           1
```

## STV_NODE_STORAGE_CAPACITY

En la tabla STV_NODE_STORAGE_CAPACITY, se muestran los detalles de la capacidad total de almacenamiento y la capacidad total utilizada de cada nodo de un clúster. Contiene una fila para cada nodo.

Solo los superusuarios pueden ver STV_NODE_STORAGE_CAPACITY. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| node | entero | El número de nodo. |
| used | entero | La cantidad de bloques de disco de 1 MB que están actualmente en uso en el nodo. Para los tipos de nodo RA3, los bloques utilizados incluyen tanto los bloques almacenados en la memoria caché local como los bloques persistentes en Amazon S3. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| capacity | entero | La capacidad total de almacenamiento aprovisionada para el nodo en bloques de 1 MB. La capacidad incluye espacio reservado por Amazon Redshift en tipos de nodos DC2 para uso interno. La capacidad es mayor que la capacidad nominal del nodo, que es la cantidad de espacio del nodo disponible para los datos del usuario. Para los tipos de nodo RA3, esta capacidad es la misma que la cuota total de almacenamiento administrado para el clúster. Para obtener más información acerca de la capacidad según el tipo de nodo, consulte Detalles acerca de los tipos de nodos en la Guía de administración de Amazon Redshift. |

Consultas de ejemplo

> **Note**
>
> Los resultados de los siguientes ejemplos varían en función de las especificaciones de los nodos de su clúster. Agregue la columna `capacity` a su SQL SELECT para recuperar la capacidad del clúster.

La siguiente consulta devuelve el espacio utilizado y la capacidad total en bloques de disco de 1 MB. Este ejemplo se ejecutó en un clúster de dos nodos dc2.8xlarge.

```
select node, used from stv_node_storage_capacity order by node;
```

Esta consulta devuelve el siguiente ejemplo de salida.

```
 node | used
------+-------
```

```
    0 | 30597
    1 | 27089
```

La siguiente consulta devuelve el espacio utilizado y la capacidad total en bloques de disco de 1 MB. Este ejemplo se ejecutó en un clúster de dos nodos ra3.16xlarge.

```
select node, used from stv_node_storage_capacity order by node;
```

Esta consulta devuelve el siguiente ejemplo de salida.

```
 node | used
------+-------
    0 | 30591
    1 | 27103
```

## STV_PARTITIONS

Utilice la tabla STV_PARTITIONS para conocer el rendimiento de la velocidad del disco y la utilización del disco para Amazon Redshift.

STV_PARTITIONS contiene una fila por nodo y por volumen lógico de disco.

Solo los superusuarios pueden ver STV_PARTITIONS. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| owner | entero | Nodo del disco que posee la partición. |
| host | entero | Nodo que está conectado físicamente a la partición. |
| diskno | entero | Disco que tiene la partición. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| part_begin | bigint | Desplazamiento de la partición. Los dispositivos en bruto están particionados de manera lógica para habilitar espacio para los bloques invertidos. |
| part_end | bigint | Fin de la partición. |
| used | entero | Cantidad de bloques de disco de 1 MB que están actualmente en uso en la partición. |
| tossed | entero | Cantidad de bloques que están listos para eliminarse, pero que no se borran porque no es seguro liberar sus direcciones de disco. Si las direcciones se liberaran de inmediato, se podría guardar una transacción pendiente en la misma ubicación en el disco. Por lo tanto, estos bloques desechados se liberan a partir de la próxima confirmación. Los bloques del disco pueden estar marcados como desechados, por ejemplo, cuando se elimina la columna de una tabla durante las operaciones INSERT o durante las operaciones de consultas basadas en el disco. |
| capacity | entero | Capacidad total de la partición en bloques de disco de 1 MB. |
| reads | bigint | Cantidad de lecturas que ocurrieron desde el último reinicio del clúster. |
| writes | bigint | Cantidad de escrituras que ocurrieron desde el último reinicio del clúster. |
| seek_forward | entero | Cantidad de veces que una solicitud no corresponde a la dirección posterior dada la dirección anterior de la solicitud. |
| seek_back | entero | Cantidad de veces que una solicitud no corresponde a la dirección anterior dada la dirección posterior. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| is_san | entero | Indica si la partición pertenece a la Storage Area Network (SAN, Red de área de almacenamiento). Los valores válidos son **0** (false) o **1** (true). |
| error | entero | Esta columna ya no está disponible. |
| mbps | entero | Velocidad del disco expresada en megabytes por segundo. |
| mount | character (256) | Ruta del directorio al dispositivo. |

Consulta de ejemplo

La siguiente consulta muestra el espacio de disco utilizado y la capacidad, en bloques de disco de 1 MB, y calcula la utilización del disco como un porcentaje del espacio de disco en bruto. El espacio de disco en bruto incluye el espacio que reserva Amazon Redshift para uso interno, por lo que supera su capacidad nominal, la cual corresponde a la cantidad de espacio de disco disponible para el usuario. La métrica Porcentaje de espacio de disco utilizado de la pestaña Performance (Rendimiento) de la consola de administración de Amazon Redshift indica el porcentaje de la capacidad nominal del disco que utiliza el clúster. Le recomendamos que monitoree la métrica Porcentaje de espacio de disco utilizado para mantener el uso dentro de la capacidad nominal del disco del clúster.

> **Important**
>
> Le recomendamos encarecidamente que no supere la capacidad nominal del disco de su clúster. Si bien puede ser técnicamente posible bajo ciertas circunstancias, superar la capacidad nominal de su disco disminuye la tolerancia a errores de su clúster y aumenta el riesgo de perder datos.

Este ejemplo se ejecutó en un clúster de dos nodos con seis particiones lógicas del disco por cada nodo. El espacio se utiliza de manera muy uniforme en todos los discos, donde se usa, aproximadamente, el 25% de cada disco.

```
select owner, host, diskno, used, capacity,
(used-tossed)/capacity::numeric *100 as pctused
from stv_partitions order by owner;

 owner | host | diskno |  used  | capacity | pctused
-------+------+--------+--------+----------+---------
   0   |  0   |    0   | 236480 |  949954  | 24.9
   0   |  0   |    1   | 236420 |  949954  | 24.9
   0   |  0   |    2   | 236440 |  949954  | 24.9
   0   |  1   |    2   | 235150 |  949954  | 24.8
   0   |  1   |    1   | 237100 |  949954  | 25.0
   0   |  1   |    0   | 237090 |  949954  | 25.0
   1   |  1   |    0   | 236310 |  949954  | 24.9
   1   |  1   |    1   | 236300 |  949954  | 24.9
   1   |  1   |    2   | 236320 |  949954  | 24.9
   1   |  0   |    2   | 237910 |  949954  | 25.0
   1   |  0   |    1   | 235640 |  949954  | 24.8
   1   |  0   |    0   | 235380 |  949954  | 24.8

(12 rows)
```

## STV_QUERY_METRICS

Tiene información acerca de las métricas, como la cantidad de filas procesadas, el uso de la CPU, la entrada/salida y el uso del disco para las consultas activas que se están ejecutando en colas de consultas (clases de servicios) definidas por el usuario. Para ver las métricas de las consultas que se han completado, consulte la tabla de sistema STL_QUERY_METRICS.

Las métricas de consultas se muestran en intervalos de un segundo. Por este motivo, diferentes ejecuciones de la misma consulta pueden devolver horas que presentan una pequeña diferencia. Además, es posible que los segmentos de consultas que se ejecutan en menos de 1 segundo no se registren.

STV_QUERY_METRICS hace un seguimiento y agrega métricas en el de nivel consulta, segmento y paso. Para obtener más información acerca de los segmentos y los pasos de las consultas, consulte Flujo de trabajo de planificación y ejecución de consultas. Distintas métricas (como `max_rows`, `cpu_time`, etc.) se suman en todos los sectores de un nodo. Para obtener más información acerca de los sectores de nodos, consulte Arquitectura del sistema de almacenamiento de datos.

Para determinar en qué nivel de fila se informa sobre las métricas, examine las columnas `segment` y `step_type`:

- Si tanto `segment` como `step_type` son igual a `-1`, la fila se informa sobre las métricas en el nivel consulta.
- Si `segment` no es `-1` y `step_type` es `-1`, la fila se informa sobre las métricas en el nivel segmento.
- Si tanto `segment` como `step_type` no son `-1`, la fila se informa sobre las métricas en el nivel paso.

STV_QUERY_METRICS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que ejecutó la consulta que generó la entrada. |
| service_class | entero | ID para la cola de consultas de WLM (clase de servicio). Las colas de consultas se definen en la configuración de WLM. Las métricas solo se informan para las colas definidas por el usuario. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| starttime | marca de tiempo | Hora en UTC en la que la consulta comenzó a ejecutarse, con 6 dígitos de precisión para los segundos fraccionados. Por ejemplo: `2009-06-12 11:29:19.131358` . |
| slices | entero | Cantidad de sectores para el clúster. |
| segment | entero | Número del segmento. Una consulta consta de distintos segmentos y cada segmento consta de uno o más pasos. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | Los segmentos de una consulta se ejecutan en paralelo. Cada segmento se ejecuta en un solo proceso. Si el valor del segmento es -1, los valores métricos del segmento se acumulan hasta el nivel consulta. |
| step_type | entero | Tipo de paso que se ejecutó. Para obtener una descripción de los tipos de pasos, consulte Tipos de pasos. |
| rows | bigint | Cantidad de filas procesadas por un paso. |
| max_rows | bigint | Cantidad máxima de filas de salida para un paso, acumulada s en todos los sectores. |
| cpu_time | bigint | Tiempo de CPU utilizado, en microsegundos. En el nivel segmento, el tiempo de CPU total para el segmento en todos los sectores. En el nivel consulta, la suma de los tiempos de CPU para la consulta en todos los sectores y segmentos. |
| max_cpu_time | bigint | Tiempo de CPU máximo utilizado, en microsegundos. En el nivel segmento, el tiempo de CPU máximo utilizado por el segmento en todos los sectores. En el nivel consulta, el tiempo de CPU máximo utilizado por cualquier segmento de la consulta. |
| blocks_read | bigint | Cantidad de bloques de 1 MB leídos por la consulta o por el segmento. |
| max_block s_read | bigint | Cantidad máxima de bloques de 1 MB leídos por el segmento, acumulados en todos los sectores. En el nivel segmento, la cantidad máxima de bloques de 1 MB leídos para el segmento en todos los sectores. En el nivel consulta, la cantidad máxima de bloques de 1 MB leídos por cualquier segmento de la consulta. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| run_time | bigint | Tiempo de ejecución total, sumado en todos los sectores. El tiempo de ejecución no incluye el tiempo de espera.<br><br>En el nivel segmento, el tiempo de ejecución para el segmento, sumado en todos los sectores. En el nivel consulta, el tiempo de ejecución para la consulta, sumado en todos los sectores y segmentos. Como este valor es una suma, el tiempo de ejecución no está relacionado con el tiempo de ejecución de la consulta. |
| max_run_time | bigint | El tiempo máximo transcurrido para un segmento, en microsegundos. En el nivel segmento, el tiempo máximo de ejecución para el segmento en todos los sectores. En el nivel consulta, el tiempo máximo de ejecución para cualquier segmento de la consulta. |
| max_block s_to_disk | bigint | La cantidad máxima de espacio de disco utilizado para grabar los resultados intermedios, en bloques de 1 MB. En el nivel segmento, la cantidad máxima de espacio en disco utilizada por el segmento en todos los sectores. En el nivel consulta, la cantidad máxima de espacio en disco utilizada por cualquier segmento de la consulta. |
| blocks_to_disk | bigint | La cantidad de espacio de disco utilizado por una consulta o un segmento para grabar los resultados intermedios, en bloques de 1 MB. |
| paso | entero | Paso de la consulta que se ejecutó. |
| max_query _scan_size | bigint | El tamaño máximo de los datos examinados por una consulta, en MB. En el nivel segmento, el tamaño máximo de datos examinados por el segmento en todos los sectores. En el nivel consulta, el tamaño máximo de datos examinados por cualquier segmento de la consulta. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| query_scan_size | bigint | El tamaño de los datos examinados por una consulta, en MB. |
| query_priority | entero | La prioridad de la consulta. Los valores posibles son `-1`, `0`, `1`, `2`, `3` y `4`, donde `-1` significa que no se admite la prioridad de la consulta. |
| query_queue_time | bigint | La cantidad de tiempo en microsegundos que la consulta estuvo en cola. |

Tipos de pasos

En la siguiente tabla, se muestran los tipos de pasos relevantes para usuarios de bases de datos. En la tabla no se muestran los tipos de pasos que son solo para uso interno. Si el tipo de paso es -1, la métrica no se informa a nivel paso.

| Step type (Tipo de paso) | Descripción |
|---|---|
| 1 | Examinar tabla |
| 2 | Insertar filas |
| 3 | Añadir filas |
| 6 | Paso de ordenación |
| 7 | Paso de fusión |
| 8 | Paso de distribución |
| 9 | Difundir paso de distribución |
| 10 | Combinación hash |
| 11 | Combinación de fusión |
| 12 | Paso guardar |

| Step type (Tipo de paso) | Descripción |
|---|---|
| 14 | Hash |
| 15 | Combinación de bucle anidado |
| 16 | Proyectar campos y expresiones |
| 17 | Limitar la cantidad de filas devueltas |
| 18 | Único |
| 20 | Eliminar filas |
| 26 | Limitar la cantidad de filas ordenadas devueltas |
| 29 | Calcular una función de ventana |
| 32 | Función definida por el usuario (UDF) |
| 33 | Paso único |
| 37 | Devolver filas del nodo principal al cliente |
| 38 | Devolver filas de los nodos de computación al nodo principal |
| 40 | Examen de Spectrum |

Consulta de ejemplo

Para encontrar consultas activas con un tiempo alto de CPU (más de 1 000 segundos), ejecute la siguiente consulta.

```
select query, cpu_time / 1000000 as cpu_seconds
from stv_query_metrics where segment = -1 and cpu_time > 1000000000
order by cpu_time;

query | cpu_seconds
------+------------
25775 |        9540
```

Para encontrar consultas activas con una combinación de bucle anidado que devolvieron más de un millón de filas, ejecute la siguiente consulta.

```
select query, rows
from stv_query_metrics
where step_type = 15 and rows > 1000000
order by rows;

query | rows
------+-----------
25775 | 1580225854
```

Para encontrar consultas activas que se ejecutaron por más de 60 segundos y que usaron menos de 10 segundos de tiempo de CPU, ejecute la siguiente consulta.

```
select query, run_time/1000000 as run_time_seconds
from stv_query_metrics
where segment = -1 and run_time > 60000000 and cpu_time < 10000000;

query | run_time_seconds
------+-----------------
25775 |              114
```

## STV_RECENTS

Utilice la tabla STV_RECENTS para obtener información acerca de las consultas que están activas actualmente y las que se han ejecutado recientemente en relación con una base de datos.

Todos los usuarios pueden ver STV_RECENTS. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_QUERY_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Solución de problemas con STV_RECENTS

STV_RECENTS es útil, en concreto, para determinar si una consulta o una colección de consultas se está ejecutando en ese momento o se ha realizado. También muestra el tiempo que se ha estado ejecutando una consulta. Esto es útil para hacerse una idea de qué consultas son de larga duración.

Puede unir STV_RECENTS a otras vistas del sistema, como STV_INFLIGHT, para recopilar metadatos adicionales sobre las consultas en ejecución. (Hay un ejemplo que muestra cómo hacerlo en la sección de consultas de ejemplo). También puede utilizar los registros devueltos desde esta vista junto con las características de monitoreo de la consola de Amazon Redshift para solucionar problemas en tiempo real.

Las vistas del sistema que complementan STV_RECENTS incluyen STL_QUERYTEXT, que recupera el texto de consulta para los comandos SQL, y SVV_QUERY_INFLIGHT, que une STV_INFLIGHT a STL_QUERYTEXT.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| status | character(20) | Estado de la consulta. Los valores válidos son **`Running`** y **`Done`**. |
| starttime | marca de tiempo | La hora en que comenzó la consulta. |
| duration | entero | Cantidad de microsegundos desde que comenzó la sesión. |
| user_name | character (50) | Nombre de usuario que ejecutó el proceso. |
| db_name | character (50) | Nombre de la base de datos. |
| consulta | character(600) | Texto de la consulta, de hasta 600 caracteres. Todo carácter adicional se trunca. |
| pid | entero | ID de proceso para la sesión asociada con la consulta, que siempre es -1 para las consultas que se han completado. |

Consultas de ejemplo

Para determinar cuáles son las consultas que se están ejecutando actualmente en la base de datos, ejecute la siguiente consulta:

```
select user_name, db_name, pid, query
```

```
from stv_recents
where status = 'Running';
```

En el siguiente ejemplo de salida, se muestra que hay una única consulta en ejecución en la base de datos TICKIT:

```
user_name | db_name |   pid   | query
----------+---------+---------+------------
dwuser    | tickit  |  19996  |select venuename, venueseats from
venue where venueseats > 50000 order by venueseats desc;
```

En el siguiente ejemplo, se devuelve una lista de consultas (en caso de haber alguna) que se están ejecutando o que están esperando en una cola para ejecutarse:

```
select * from stv_recents where status<>'Done';

status |    starttime        | duration |user_name|db_name| query     | pid
-------+---------------------+----------+---------+-------+-----------+------
Running| 2010-04-21 16:11... | 281566454| dwuser  |tickit | select ...| 23347
```

Esta consulta no devuelve resultados a menos que esté ejecutando una serie de consultas simultáneas y algunas estén en una cola.

En el siguiente ejemplo, se amplía el ejemplo anterior. En este caso, las consultas que están realmente "en tránsito" (ejecutándose, no esperando) no están incluidas en el resultado:

```
select * from stv_recents where status<>'Done'
and pid not in (select pid from stv_inflight);
...
```

Para obtener más consejos sobre la resolución de problemas de rendimiento de las consultas, consulte Solución de problemas de consultas.

## STV_SESSIONS

Utilice la tabla STV_SESSIONS para ver información acerca de las sesiones de usuario activas para Amazon Redshift.

Para ver el historial de sesiones, utilice la tabla STL_SESSIONS, en lugar de STV_SESSIONS.

STV_SESSIONS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_SESSION_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| starttime | marca de tiempo | Hora de inicio de la sesión. |
| procesar | entero | ID de proceso para la sesión. |
| user_name | character (50) | Usuario asociado con la sesión. |
| db_name | character (50) | Nombre de la base de datos asociada a la sesión. |
| timeout_sec | int | El tiempo máximo en segundos durante el cual una sesión permanece inactiva o en reposo antes de que se cierre. El 0 indica que no se establece tiempo de espera. |

Consultas de ejemplo

Para verificar rápidamente si hay algún otro usuario en este momento con una sesión abierta en Amazon Redshift, escriba la siguiente consulta:

```
select count(*)
from stv_sessions;
```

Si el resultado es mayor que uno, hay al menos un usuario más en este momento con una sesión abierta en la base de datos.

Para ver todas las sesiones activas de Amazon Redshift, escriba la siguiente consulta:

```
select *
```

```
from stv_sessions;
```

El siguiente resultado muestra cuatro sesiones activas actualmente en Amazon Redshift:

```
       starttime         | process |user_name                   | db_name
      | timeout_sec
--------------------------+---------+----------------------------
+---------------------------+------------
  2018-08-06 08:44:07.50 |   13779 | IAMA:aws_admin:admin_grp   | dev
      | 0
  2008-08-06 08:54:20.50 |   19829 | dwuser                     | dev
      | 120
  2008-08-06 08:56:34.50 |   20279 | dwuser                     | dev
      | 120
  2008-08-06 08:55:00.50 |   19996 | dwuser                     | tickit
      | 0
(3 rows)
```

En nombre de usuario con el prefijo IAMA indica que el usuario inició una sesión con el inicio de sesión único federado. Para obtener más información, consulte Uso de la autenticación de IAM para generar credenciales de usuario de base de datos.

## STV_SLICES

Utilice la tabla STV_SLICES para ver el mapeo actual de un sector a un nodo.

La información de STV_SLICES se utiliza, principalmente, para investigaciones.

Todos los usuarios pueden ver STV_SLICES. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| node | entero | Nodo del clúster donde se encuentra el sector. |
| slice | entero | Sector del nodo. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| localslice | entero | Esta información es solo para uso interno. |
| type | character (1) | Esta información es solo para uso interno. |

Consulta de ejemplo

Para ver qué nodos del clúster está administrando cada sector, escriba la siguiente consulta:

```
select node, slice from stv_slices;
```

Esta consulta devuelve el siguiente ejemplo de salida:

```
 node | slice
------+-------
    0 |     2
    0 |     3
    0 |     1
    0 |     0
(4 rows)
```

## STV_STARTUP_RECOVERY_STATE

Registra el estado de las tablas que están bloqueadas temporalmente durante las operaciones de reinicio del clúster. Amazon Redshift bloquea de manera temporal las tablas mientras estas se procesan para resolver transacciones obsoletas después de que se reinicia un clúster.

STV_STARTUP_RECOVERY_STATE es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| db_id | entero | ID de la base de datos. |
| table_id | entero | ID de la tabla. |
| table_name | character(137) | Nombre de la tabla. |

Consultas de ejemplo

Para monitorear qué tablas están bloqueadas temporalmente, ejecute la siguiente consulta después de que se reinicie un clúster.

```
select * from STV_STARTUP_RECOVERY_STATE;

  db_id | tbl_id | table_name
--------+--------+------------
 100044 | 100058 | lineorder
 100044 | 100068 | part
 100044 | 100072 | customer
 100044 | 100192 | supplier
(4 rows)
```

## STV_TBL_PERM

La tabla STV_TBL_PERM tiene información acerca de las tablas permanentes en Amazon Redshift, incluidas las tablas temporales creadas por un usuario para esta sesión. STV_TBL_PERM tiene información de todas las tablas en todas las bases de datos.

Esta tabla se diferencia de STV_TBL_TRANS, la cual tiene información relacionada con las tablas transitorias de bases de datos que el sistema crea durante el procesamiento de consultas.

Solo los superusuarios pueden ver STV_TBL_PERM. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| slice | entero | Sector del nodo asignado a la tabla. |
| id | entero | ID de la tabla. |
| nombre | character (72) | Nombre de la tabla. |
| rows | bigint | Cantidad de fila de datos en el sector. |
| sorted_rows | bigint | Cantidad de filas en el sector que ya están ordenadas en el disco. Si esta cantidad no coincide con la cantidad indicada en ROWS, limpie la tabla para volver a ordenar las filas. |
| temp | entero | Indica si es una tabla temporal o no: 0 = false; 1 = true. |
| db_id | entero | ID de la base de datos donde se crea la tabla. |
| insert_pristine | entero | Para uso interno. |
| delete_pristine | entero | Para uso interno. |
| copia de seguridad | entero | Valor que indica si la tabla se incluye en las instantáneas del clúster: 0 = no; 1 = sí. Para obtener más información, consulte el parámetro BACKUP del comando CREATE TABLE. |
| dist_style | entero | Estilo de distribución de la tabla a la que pertenece el segmento. Para obtener más información acerca de los valores, consulte Visualización de los estilos de distribución. Para obtener más información acerca de los estilos de distribución, consulte Estilos de distribución. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| block_count | entero | Número de bloques que utiliza el segmento. El valor es -1 cuando no se puede calcular el número de bloques. |

Consultas de ejemplo

La siguiente consulta devuelve una lista de diferentes ID y nombres de tablas:

```
select distinct id, name
from stv_tbl_perm order by name;

   id   |          name
--------+------------------------
 100571 | category
 100575 | date
 100580 | event
 100596 | listing
 100003 | padb_config_harvest
 100612 | sales
...
```

Otras tablas de sistema usan ID de tablas, por lo que puede resultar muy útil saber qué ID de tabla corresponde a una tabla determinada. En este ejemplo, se utiliza el comando SELECT DISTINCT para eliminar duplicados (las tablas se distribuyen en diferente sectores).

Para determinar la cantidad de bloques que utiliza cada columna de la tabla VENUE, escriba la siguiente consulta:

```
select col, count(*)
from stv_blocklist, stv_tbl_perm
where stv_blocklist.tbl = stv_tbl_perm.id
and stv_blocklist.slice = stv_tbl_perm.slice
and stv_tbl_perm.name = 'venue'
group by col
order by col;

 col | count
```

```
-----+-------
   0 |     8
   1 |     8
   2 |     8
   3 |     8
   4 |     8
   5 |     8
   6 |     8
   7 |     8
(8 rows)
```

Notas de uso

La columna ROWS incluye los recuentos de las filas eliminadas que no se limpiaron (o que se limpiaron, pero con la opción SORT ONLY). Por lo tanto, es posible que el resultado de SUM de la columna ROWS en la tabla STV_TBL_PERM no coincida con el resultado de COUNT(*) cuando consulta directamente una tabla determinada. Por ejemplo, si se eliminan 2 filas de VENUE, el resultado de COUNT(*) es 200, pero el resultado de SUM(ROWS) sigue siendo 202:

```
delete from venue
where venueid in (1,2);

select count(*) from venue;
count
-------
200
(1 row)

select trim(name) tablename, sum(rows)
from stv_tbl_perm where name='venue' group by name;

tablename | sum
-----------+-----
venue     | 202
(1 row)
```

Para sincronizar los datos en STV_TBL_PERM, ejecute una limpieza completa en la tabla VENUE.

```
vacuum venue;

select trim(name) tablename, sum(rows)
from stv_tbl_perm
```

```
where name='venue'
group by name;

tablename | sum
-----------+-----
venue     | 200
(1 row)
```

## STV_TBL_TRANS

Utilice la tabla STV_TBL_TRANS para encontrar información acerca de las tablas transitorias de las bases de datos que están actualmente en la memoria.

Las tablas transitorias son, por lo general, conjuntos de filas temporales que se usan como resultados intermedios mientras se ejecuta una consulta. STV_TBL_TRANS se diferencia de STV_TBL_PERM debido a que STV_TBL_PERM tiene información acerca de las tablas permanentes de las bases de datos.

Solo los superusuarios pueden ver STV_TBL_TRANS. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| slice | entero | Sector del nodo asignado a la tabla. |
| id | entero | ID de la tabla. |
| rows | bigint | Cantidad de fila de datos en la tabla. |
| size | bigint | Cantidad de bytes asignados a la tabla. |
| query_id | bigint | ID de la consulta. |
| ref_cnt | entero | Cantidad de referencias. |
| from_suspended | entero | Indica si esta tabla fue creada o no durante una consulta que ahora está suspendida. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| prep_swap | entero | Indica si una tabla transitoria está o no preparada para cambiarse al disco si fuera necesario. (El cambio solo ocurrirá en situaciones en las que la memoria es baja). |

Consultas de ejemplo

Para ver información acerca de las tablas transitorias para una consulta con un ID de consulta de 90, escriba el siguiente comando:

```
select slice, id, rows, size, query_id, ref_cnt
from stv_tbl_trans
where query_id = 90;
```

Esta consulta devuelve información acerca de las tablas transitorias para una consulta 90, tal como se muestra en el siguiente ejemplo de salida:

```
slice | id | rows | size | query_ | ref_ | from_     | prep_
      |    |      |      | id     | cnt  | suspended | swap
------+----+------+------+--------+------+-----------+-------
 1013 | 95 |    0 |    0 |     90 |    4 |         0 |   0
    7 | 96 |    0 |    0 |     90 |    4 |         0 |   0
   10 | 96 |    0 |    0 |     90 |    4 |         0 |   0
   17 | 96 |    0 |    0 |     90 |    4 |         0 |   0
   14 | 96 |    0 |    0 |     90 |    4 |         0 |   0
    3 | 96 |    0 |    0 |     90 |    4 |         0 |   0
 1013 | 99 |    0 |    0 |     90 |    4 |         0 |   0
    9 | 96 |    0 |    0 |     90 |    4 |         0 |   0
    5 | 96 |    0 |    0 |     90 |    4 |         0 |   0
   19 | 96 |    0 |    0 |     90 |    4 |         0 |   0
    2 | 96 |    0 |    0 |     90 |    4 |         0 |   0
 1013 | 98 |    0 |    0 |     90 |    4 |         0 |   0
   13 | 96 |    0 |    0 |     90 |    4 |         0 |   0
    1 | 96 |    0 |    0 |     90 |    4 |         0 |   0
 1013 | 96 |    0 |    0 |     90 |    4 |         0 |   0
    6 | 96 |    0 |    0 |     90 |    4 |         0 |   0
   11 | 96 |    0 |    0 |     90 |    4 |         0 |   0
   15 | 96 |    0 |    0 |     90 |    4 |         0 |   0
```

```
   18 | 96 |    0 |    0 |     90 |    4 |         0 |   0
```

En este ejemplo, puede ver que los datos de la consulta involucran a las tablas 95, 96 y 98. Como no hay ningún byte asignado a esta tabla, esta consulta puede ejecutarse en la memoria.

## STV_WLM_CLASSIFICATION_CONFIG

Tiene las reglas actuales de clasificación para WLM.

Solo los superusuarios pueden ver STV_WLM_CLASSIFICATION_CONFIG. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| id | entero | ID de la clase de servicio. Para obtener una lista de ID de clase de servicio, consulte ID de la clase de servicio WLM. |
| condition | character(128) | Condiciones de la consulta. |
| action_seq | entero | Reservado para el uso del sistema. |
| acción | character(64) | Reservado para el uso del sistema. |
| action_service_class | entero | La clase de servicios donde ocurre la acción. |

Consulta de ejemplo

```
select * from STV_WLM_CLASSIFICATION_CONFIG;

id | condition                                    | action_seq | action |
 action_service_class
---+----------------------------------------------+------------+--------
+---------------------
 1 | (system user) and (query group: health)     |          0 | assign |
     1
 2 | (system user) and (query group: metrics)    |          0 | assign |
     2
```

```
 3 | (system user) and (query group: cmstats)    |          0 | assign |
     3
 4 | (system user)                               |          0 | assign |
     4
 5 | (super user) and (query group: superuser)   |          0 | assign |
     5
 6 | (query group: querygroup1)                  |          0 | assign |
     6
 7 | (user group: usergroup1)                    |          0 | assign |
     6
 8 | (user group: usergroup2)                    |          0 | assign |
     7
 9 | (query group: querygroup3)                  |          0 | assign |
     8
10 | (query group: querygroup4)                  |          0 | assign |
     9
11 | (user group: usergroup4)                    |          0 | assign |
     9
12 | (query group: querygroup*)                  |          0 | assign |
    10
13 | (user group: usergroup*)                    |          0 | assign |
    10
14 | (querytype: any)                            |          0 | assign |
    11
(4 rows)
```

## STV_WLM_QMR_CONFIG

Registra la configuración para las Query Monitoring Rules (QMR, Reglas de monitorización de consultas) de WLM. Para obtener más información, consulte Reglas de monitoreo de consultas de WLM.

Solo los superusuarios pueden ver STV_WLM_QMR_CONFIG. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| service_class | entero | ID para la cola de consultas de WLM (clase de servicio) . Las colas de consultas se definen en la configuración |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | de WLM. Las reglas solo pueden definirse para las colas definidas por el usuario. Para obtener una lista de ID de clase de servicio, consulte ID de la clase de servicio WLM. |
| rule_name | character (256) | Nombre de la regla de monitorización de consultas. |
| acción | character (256) | Acción de la regla. Los valores posibles son `log`, `hop`, `abort` y `change_query_priority` . |
| metric_name | character (256) | Nombre de la métrica. |
| metric_operator | character (256) | El operador de la métrica. Los valores posibles son >, =, <. |
| metric_value | double | El valor de umbral para la métrica especificada que activa una acción. |
| action_value | character (256) | Si `action` es `change_query_priority` , los valores posibles son `highest`, `high`, `normal`, `low` y `lowest`.<br><br>Si `action` es `log`, hop o `abort`, el valor está vacío. |

Consulta de ejemplo

Para ver las definiciones de las reglas de QMR para todas las clases de servicio mayores que 5 (se incluyen colas definidas por el usuario), ejecute la siguiente consulta. Para obtener una lista de ID de clase de servicio, consulte ID de la clase de servicio WLM.

```
Select *
from stv_wlm_qmr_config
where service_class > 5
order by service_class;
```

## STV_WLM_QUERY_QUEUE_STATE

Registra el estado actual de las colas de consultas para las clases de servicios.

Todos los usuarios pueden ver STV_WLM_QUERY_QUEUE_STATE. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_QUERY_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| service_class | entero | ID para la clase de servicio. Para obtener una lista de ID de clase de servicio, consulte ID de la clase de servicio WLM. |
| position | entero | Posición de la consulta en la cola. La consulta que tiene el valor de **`position`** más pequeño se ejecuta a continuación. |
| tarea | entero | ID utilizado para hacer un seguimiento de una consulta a través del administrador de cargas de trabajo. Puede estar asociado a distintos ID de consultas. Si se reinicia una consulta, esta se asigna a un ID de consulta nuevo, pero no a un ID de tarea nuevo. |
| consulta | entero | ID de la consulta. Si se reinicia una consulta, esta se asigna a un ID de consulta nuevo, pero no a un ID de tarea nuevo. |
| slot_count | entero | Cantidad de slots de consulta de WLM. |
| start_time | marca de tiempo | Hora en la que ingresó la consulta a la cola. |
| queue_time | bigint | Cantidad de microsegundos que estuvo la consulta en la cola. |

Consulta de ejemplo

La siguiente consulta muestra las consultas en la cola para las clases de servicio mayores que 4.

```
select * from stv_wlm_query_queue_state
where service_class > 4
order by service_class;
```

Esta consulta devuelve el siguiente ejemplo de salida.

```
 service_class | position | task | query | slot_count |        start_time          |
 queue_time
---------------+----------+------+-------+------------+----------------------------
+------------
             5 |        0 |  455 |   476 |          5 | 2010-10-06 13:18:24.065838 |
 20937257
             6 |        1 |  456 |   478 |          5 | 2010-10-06 13:18:26.652906 |
 18350191
(2 rows)
```

## STV_WLM_QUERY_STATE

Registra el estado actual de las consultas de las que WLM hace un seguimiento.

Todos los usuarios pueden ver STV_WLM_QUERY_STATE. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_QUERY_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| xid | entero | ID de la transacción de la consulta o de la subconsulta. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| tarea | entero | ID utilizado para hacer un seguimiento de una consulta a través del administrador de cargas de trabajo. Puede estar asociado a distintos ID de consultas. Si se reinicia una consulta, esta se asigna a un ID de consulta nuevo, pero no a un ID de tarea nuevo. |
| consulta | entero | ID de la consulta. Si se reinicia una consulta, esta se asigna a un ID de consulta nuevo, pero no a un ID de tarea nuevo. |
| service_class | entero | ID para la clase de servicio. Para obtener una lista de ID de clase de servicio, consulte ID de la clase de servicio WLM. |
| slot_count | entero | Cantidad de slots de consulta de WLM. |
| wlm_start_time | marca de tiempo | Hora en la que la consulta ingresó a la cola de las tablas de sistema o a la cola de consultas cortas. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| estado | character(16) | Estado actual de la consulta o de la subconsulta.<br><br>Los valores posibles son los siguientes:<br><br>• `Classified` : la consulta se ha asignado a una clase de servicio.<br>• `Completed` : la consulta ha terminado de ejecutars e. La consulta se ha ejecutado correctamente o se ha cancelado. Para conocer el estado final, verifique los resultados de STL_QUERY.<br>• `Dequeued`: solo uso interno.<br>• `Evicted`: se ha expulsado la consulta de la clase de servicio para reiniciarla.<br>• `Evicting`: se está expulsando la consulta de la clase de servicio para reiniciarla.<br>• `Initialized` : solo uso interno.<br>• `Invalid`: solo uso interno.<br>• `Queued`: la consulta se envió a la cola de consultas porque no había secciones disponibles para ejecutarla.<br>• `QueuedWaiting` : la consulta está esperando en la cola de consultas.<br>• `Rejected`: solo uso interno.<br>• `Returning` : la consulta está devolviendo los resultado s al cliente.<br>• `Running`: la consulta se está ejecutando.<br>• `TaskAssigned` : solo uso interno. |
| queue_tim e | bigint | Cantidad de microsegundos que estuvo la consulta en la cola. |
| exec_time | bigint | Cantidad de microsegundos que la consulta estuvo en ejecución. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| query_priority | char(20) | La prioridad de la consulta. Los valores posibles son n/a, `lowest`, `low`, `normal`, `high` y `highest`, donde n/a significa que no se admite la prioridad de la consulta. |

Consulta de ejemplo

La siguiente consulta muestra todas las consultas que se están ejecutando actualmente en las clases de servicio mayores que 4. Para obtener una lista de ID de clase de servicio, consulte ID de la clase de servicio WLM.

```
select xid, query, trim(state) as state, queue_time, exec_time
from stv_wlm_query_state
where service_class > 4;
```

Esta consulta devuelve el siguiente ejemplo de salida:

```
xid    | query | state   | queue_time | exec_time
-------+-------+---------+------------+-----------
100813 | 25942 | Running |          0 |    1369029
100074 | 25775 | Running |          0 | 2221589242
```

## STV_WLM_QUERY_TASK_STATE

Tiene el estado actual de las tareas de consultas de las clases de servicios.

Todos los usuarios pueden ver STV_WLM_QUERY_TASK_STATE. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| service_class | entero | ID para la clase de servicio. Para obtener una lista de ID de clase de servicio, consulte ID de la clase de servicio WLM. |
| tarea | entero | ID utilizado para hacer un seguimiento de una consulta a través del administrador de cargas de trabajo. Puede estar asociado a distintos ID de consultas. Si se reinicia una consulta, esta se asigna a un ID de consulta nuevo, pero no a un ID de tarea nuevo. |
| consulta | entero | ID de la consulta. Si se reinicia una consulta, esta se asigna a un ID de consulta nuevo, pero no a un ID de tarea nuevo. |
| slot_count | entero | Cantidad de slots de consulta de WLM. |
| start_time | marca de tiempo | Hora en la se comenzó a ejecutar la consulta. |
| exec_time | bigint | Cantidad de microsegundos que la consulta estuvo en ejecución. |

Consulta de ejemplo

La siguiente consulta muestra el estado actual de las consultas en las clases de servicio mayores que 4. Para obtener una lista de ID de clase de servicio, consulte ID de la clase de servicio WLM.

```
select * from stv_wlm_query_task_state
where service_class > 4;
```

Esta consulta devuelve el siguiente ejemplo de salida:

```
service_class | task | query |         start_time         | exec_time
--------------+------+-------+----------------------------+-----------
```

```
    5         |  466 |   491 | 2010-10-06 13:29:23.063787 | 357618748
(1 row)
```

## STV_WLM_SERVICE_CLASS_CONFIG

Registra las configuraciones de las clases de servicios para WLM.

Solo los superusuarios pueden ver STV_WLM_SERVICE_CLASS_CONFIG. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| service_class | entero | ID para la clase de servicio. Para obtener una lista de ID de clase de servicio, consulte ID de la clase de servicio WLM. |
| queueing_strategy | character(32) | Reservado para el uso del sistema. |
| num_query_tasks | entero | Nivel de simultaneidad real actual de la clase de servicio. Si `num_query_tasks` y `target_num_query_tasks` son diferentes, hay una transición dinámica de WLM en proceso. Un valor de `-1` indica que Auto WLM (WLM automática) se ha configurado. |
| target_num_query_tasks | entero | Nivel de simultaneidad configurado por el cambio de configuración de WLM más reciente. |
| evictable | character(8) | Reservado para el uso del sistema. |
| eviction_threshold | bigint | Reservado para el uso del sistema. |
| query_working_mem | entero | Cantidad real actual de memoria funcional, en MB por slot y por nodo, asignada a la clase de servicio. Si `query_working_mem` y `target_query_working_mem` son diferentes, hay una transición dinámica de WLM en proceso. Un valor de `-1` indica que Auto WLM (WLM automática) se ha configurado. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| target_query_working_mem | entero | Cantidad de memoria funcional, en MB por slot y por nodo, configurada por el cambio de configuración de WLM más reciente. |
| min_step_mem | entero | Reservado para el uso del sistema. |
| nombre | character(64) | El nombre de la clase de servicio. |
| max_execution_time | bigint | Cantidad de milisegundos que la consulta puede estar en ejecución antes de finalizar. |
| user_group_wild_card | Booleano | Si TRUE, la cola de WLM considera al asterisco (*) como un carácter comodín en cadenas de grupos de usuarios en la configuración de WLM. |
| query_group_wild_card | Booleano | Si TRUE, la cola de WLM considera al asterisco (*) como un carácter comodín en cadenas de grupos de consultas en la configuración de WLM. |
| concurrency_scaling | character(20) | Describe si el escalado de simultaneidad se encuentra como `on` o `off`. |
| query_priority | character(20) | El valor en términos de prioridad de la consulta. |
| user_role_wild_card | Booleano | Si TRUE, la cola de WLM considera al asterisco (*) como un carácter comodín en cadenas de usuarios en la configuración de WLM. |

Consulta de ejemplo

La primera clase de servicio definida por el usuario es la clase de servicio 6, que se denominada clase de servicio n. ° 1. La siguiente consulta muestra la configuración actual de las clases de servicio mayores que 4. Para obtener una lista de ID de clase de servicio, consulte ID de la clase de servicio WLM.

```
select rtrim(name) as name,
```

```
num_query_tasks as slots,
query_working_mem as mem,
max_execution_time as max_time,
user_group_wild_card as user_wildcard,
query_group_wild_card as query_wildcard
from stv_wlm_service_class_config
where service_class > 4;

name                          | slots | mem | max_time | user_wildcard | query_wildcard
------------------------------+-------+-----+----------+---------------+---------------
Service class for super user |     1 | 535 |        0 | false         | false
Queue 1                       |     5 | 125 |        0 | false         | false
Queue 2                       |     5 | 125 |        0 | false         | false
Queue 3                       |     5 | 125 |        0 | false         | false
Queue 4                       |     5 | 627 |        0 | false         | false
Queue 5                       |     5 | 125 |        0 | true          | true
Default queue                 |     5 | 125 |        0 | false         | false
```

La siguiente consulta muestra el estado de una transición dinámica de WLM. Mientras la transición está en proceso, `num_query_tasks` y `target_query_working_mem` se actualizan hasta que se equiparan con los valores de destino. Para obtener más información, consulte Propiedades de configuración dinámicas y estáticas de WLM.

```
select rtrim(name) as name,
num_query_tasks as slots,
target_num_query_tasks as target_slots,
query_working_mem as memory,
target_query_working_mem as target_memory
from stv_wlm_service_class_config
where num_query_tasks > target_num_query_tasks
or query_working_mem > target_query_working_mem
and service_class > 5;

 name             | slots | target_slots | memory | target_mem
------------------+-------+--------------+--------+------------
 Queue 3          |     5 |           15 |    125 |       375
 Queue 5          |    10 |            5 |    250 |       125
 (2 rows)
```

## STV_WLM_SERVICE_CLASS_STATE

Tiene el estado actual de las clases de servicios.

Solo los superusuarios pueden ver STV_WLM_SERVICE_CLASS_STATE. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| service_class | entero | ID para la clase de servicio. Para obtener una lista de ID de clase de servicio, consulte ID de la clase de servicio WLM. |
| num_queued_queries | entero | Cantidad de consultas que están actualmente en la cola. |
| num_executing_queries | entero | Cantidad de consultas que se están ejecutando actualmente. |
| num_serviced_queries | entero | Cantidad de consultas que estuvieron alguna vez en la clase de servicio. |
| num_executed_queries | entero | Cantidad de consultas que se ejecutaron desde que se reinició Amazon Redshift. |
| num_evicted_queries | entero | Cantidad de consultas que se expulsaron desde que se reinició Amazon Redshift. Algunos de los motivos para descartar una consulta incluyen: tiempo de espera de WLM agotado, una acción de salto de QMR o una consulta que ha producido un error en un clúster de escalado de simultaneidad. |
| num_concurrency_scaling_queries | entero | Cantidad de consultas que se ejecutaron en un clúster de escalado de simultaneidad desde que se reinició Amazon Redshift. |

Consulta de ejemplo

La siguiente consulta muestra el estado de las clases de servicio mayores que 5. Para obtener una lista de ID de clase de servicio, consulte ID de la clase de servicio WLM.

```
select service_class, num_executing_queries,
```

```
num_executed_queries
from stv_wlm_service_class_state
where service_class > 5
order by service_class;
```

```
 service_class | num_executing_queries | num_executed_queries
---------------+-----------------------+----------------------
             6 |                     1 |                  222
             7 |                     0 |                  135
             8 |                     1 |                   39
(3 rows)
```

## STV_XRESTORE_ALTER_QUEUE_STATE

Utilice STV_XRESTORE_ALTER_QUEUE_STATE para monitorear el progreso de la migración de cada tabla durante un cambio de tamaño clásico. Esto es aplicable específicamente cuando el tipo de nodo de destino es RA3. Para obtener más información sobre el cambio de tamaño clásico en los nodos RA3, consulte Cambio de tamaño clásico.

Solo los superusuarios pueden ver STV_XRESTORE_ALTER_QUEUE_STATE. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_RESTORE_STATE. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que inició el cambio de tamaño. |
| db_id | entero | ID de la base de datos. |
| esquema | char(128) | El nombre del esquema. |
| table_name | char(128) | El nombre de la tabla. |
| tbl | entero | ID de la tabla. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| status | char(64) | Estado del progreso de la migración de la tabla. Los valores posibles son los siguientes.<br>• `Waiting`: esperando a que comience la redistribución<br>• `Applying`: se está redistribuyendo en ese momento<br>• `Finished`: se ha terminado de redistribuir |
| task_type | entero | Es el tipo de redistribución de la tabla. Los valores posibles son los siguientes.<br>• 1: KEY<br>• 2: EVEN<br><br>Para obtener más información acerca de los estilos de distribución, consulte Estilos de distribución. |

Consulta de ejemplo

La siguiente consulta muestra el número de tablas de una base de datos que están esperando a que se cambien de tamaño, que se están cambiando de tamaño en ese momento y que se han terminado de cambiar de tamaño.

```
select db_id, status, count(*)
from stv_xrestore_alter_queue_state
group by 1,2 order by 3 desc

db_id  |   status    | count
-------+-------------+------
694325 | Waiting     |   323
694325 | Finished    |    60
694325 | Applying    |     1
```

# Vistas de SVCS para clústeres de escalado principales y de simultaneidad

Las vistas del sistema SVCS con el prefijo SVCS proporcionan información sobre las consultas tanto de los clústeres principales como de los clústeres de escalado de simultaneidad. Las vistas son similares a las tablas con el prefijo STL, con la salvedad de que las tablas STL solo proporcionan información para las consultas ejecutadas en el clúster principal.

Temas


- SVCS_ALERT_EVENT_LOG
- SVCS_COMPILE
- SVCS_CONCURRENCY_SCALING_USAGE
- SVCS_EXPLAIN
- SVCS_PLAN_INFO
- SVCS_QUERY_SUMMARY
- SVCS_S3LIST
- SVCS_S3LOG
- SVCS_S3PARTITION_SUMMARY
- SVCS_S3QUERY_SUMMARY
- SVCS_STREAM_SEGS
- SVCS_UNLOAD_LOG


## SVCS_ALERT_EVENT_LOG

Registra una alerta cuando el optimizador de consultas identifica condiciones que podrían indicar problemas de rendimiento. Esta vista se obtiene de la tabla del sistema STL_ALERT_EVENT_LOG, pero no muestra el sector de las consultas ejecutadas en un clúster de escalado de simultaneidad. Utilice la tabla SVCS_ALERT_EVENT_LOG para identificar oportunidades de mejora del rendimiento de las consultas.

Una consulta consta de distintos segmentos y cada segmento consta de uno o más pasos. Para obtener más información, consulte Procesamiento de consultas.

> **Note**
>
> Las vistas del sistema con el prefijo SVCS proporcionan información sobre las consultas de los clústeres principal y de escalado de simultaneidad. Las vistas son similares a las tablas con el prefijo STL, con la salvedad de que las tablas STL solo proporcionan información para las consultas ejecutadas en el clúster principal.

Todos los usuarios pueden ver SVCS_ALERT_EVENT_LOG. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| userid | integer | ID del usuario que generó la entrada. |
| consulta | integer | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| segment | integer | Número que identifica el segmento de la consulta. |
| paso | integer | Paso de la consulta que se ejecutó. |
| pid | integer | ID de proceso asociado a la instrucción y al sector. La misma consulta podría tener varios PID si se ejecutara en varios sectores. |
| xid | bigint | ID de la transacción asociado a la instrucción. |
| evento | character (1024) | Descripción del evento de alerta. |
| solution | character (1024) | Solución recomendada. |

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| event_time | Marca de tiempo | Hora en UTC en que comenzó la consulta. El tiempo total incluye colas y ejecución. Con 6 dígitos de precisión para fracciones de segundo. Por ejemplo: **`2009-06-12 11:29:19.131358`** . |

Notas de uso

Puede utilizar SVCS_ALERT_EVENT_LOG para identificar posibles problemas en sus consultas; luego, siga las prácticas descritas en Ajuste del rendimiento de las consultas para optimizar el diseño de su base de datos y vuelva a escribir sus consultas. SVCS_ALERT_EVENT_LOG registra las siguientes alertas:

- Faltan estadísticas

  Faltan estadísticas. Ejecute la función ANALYZE después de la carga de datos o de actualizaciones importantes y utilice la función STATUPDATE con las operaciones COPY. Para obtener más información, consulte Prácticas recomendadas de Amazon Redshift para el diseño de consultas.

- Bucle anidado

  Un bucle anidado es, por lo general, un producto cartesiano. Evalúe su consulta para asegurarse de que todas las tablas que emplee estén combinadas de forma eficaz.

- Filtro muy selectivo

  La relación de filas devueltas respecto a las filas examinadas es inferior al 0,05. Las filas escaneadas se obtienen del valor de `rows_pre_user_filter` y las filas devueltas, del valor de las filas de la tabla de sistema STL_SCAN. Indica que la consulta está examinando una cantidad excepcionalmente grande de filas para determinar el conjunto de resultados. Esto puede deberse a que faltan claves de ordenación, o que estas son incorrectas. Para obtener más información, consulte Claves de clasificación.

- Filas fantasma excesivas

  Un examen omitió una cantidad considerablemente grande de filas que están marcadas como eliminadas pero no limpiadas o filas que fueron insertadas pero no confirmadas. Para obtener más información, consulte Limpieza de tablas.

- Distribución grande

  Más de 1 000 000 filas fueron redistribuidas para una combinación hash o una agregación. Para obtener más información, consulte Distribución de datos para la optimización de consultas.

- Difusión grande

  Más de 1 000 000 filas fueron difundidas para una combinación hash. Para obtener más información, consulte Distribución de datos para la optimización de consultas.

- Ejecución en serie

  En el plan de consulta se indicó un estilo de redistribución DS_DIST_ALL_INNER, lo que exige una ejecución en serie porque toda la tabla interna fue redistribuida a un nodo único. Para obtener más información, consulte Distribución de datos para la optimización de consultas.

### Consultas de ejemplo

La siguiente consulta muestra eventos de alerta para cuatro consultas.

```
SELECT query, substring(event,0,25) as event,
substring(solution,0,25) as solution,
trim(event_time) as event_time from svcs_alert_event_log order by query;

 query |             event             |          solution            |     event_time

-------+-------------------------------+------------------------------
+---------------------
  6567 | Missing query planner statist | Run the ANALYZE command      | 2014-01-03
 18:20:58
  7450 | Scanned a large number of del | Run the VACUUM command to rec| 2014-01-03
 21:19:31
  8406 | Nested Loop Join in the query | Review the join predicates to| 2014-01-04
 00:34:22
 29512 | Very selective query filter:r | Review the choice of sort key| 2014-01-06
 22:00:00

(4 rows)
```

## SVCS_COMPILE

Registra el tiempo de compilación y la ubicación de cada segmento de consulta, incluidas las consultas ejecutadas en un clúster de escalado y las ejecutadas en el clúster principal.

> **Note**
>
> Las vistas del sistema con el prefijo SVCS proporcionan información sobre las consultas de los clústeres principal y de escalado de simultaneidad. Las vistas son similares a las vista con el prefijo SVL, con la salvedad de que las vistas SVL solo proporcionan información para las consultas ejecutadas en el clúster principal.

SVCS_COMPILE es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Para obtener información sobre SCL_COMPILE, consulte SVL_COMPILE.

Columnas de la tabla

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| userid | integer | ID del usuario que generó la entrada. |
| xid | bigint | El ID de la transacción asociado a la instrucción. |
| pid | integer | El ID del proceso asociado a la instrucción. |
| consulta | integer | El ID de la consulta. Puede usar este ID para combinar distintas vistas y tablas del sistema. |
| segment | integer | El segmento de la consulta que se compilará. |
| locus | integer | La ubicación en la que se ejecuta el segmento, **1** si es en un nodo de informática y **2** si es en el nodo principal. |
| starttime | Marca de tiempo | La hora en Tiempo coordinado universal (UTC) a la que comenzó la compilación. |

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| endtime | Marca de tiempo | La hora en UTC en que finalizó la compilación. |
| compile | integer | Un valor que es **0** si la compilación se reutilizó y **1** si se compiló el segmento. |

Consultas de ejemplo

En este ejemplo, las consultas 35878 y 35879 ejecutaron la misma instrucción SQL. La columna de compilación para la consulta 35878 muestra 1 para cuatro segmentos de consulta, lo que indica que esos segmentos se compilaron. La consulta 35879 muestra `0` en la columna de compilación para cada segmento, lo que indica que esos segmentos no se volvieron a compilar.

```
select userid, xid,  pid, query, segment, locus,
datediff(ms, starttime, endtime) as duration, compile
from svcs_compile
where query = 35878 or query = 35879
order by query, segment;

 userid |  xid   |  pid  | query | segment | locus | duration | compile
--------+--------+-------+-------+---------+-------+----------+---------
    100 | 112780 | 23028 | 35878 |       0 |     1 |        0 |       0
    100 | 112780 | 23028 | 35878 |       1 |     1 |        0 |       0
    100 | 112780 | 23028 | 35878 |       2 |     1 |        0 |       0
    100 | 112780 | 23028 | 35878 |       3 |     1 |        0 |       0
    100 | 112780 | 23028 | 35878 |       4 |     1 |        0 |       0
    100 | 112780 | 23028 | 35878 |       5 |     1 |        0 |       0
    100 | 112780 | 23028 | 35878 |       6 |     1 |     1380 |       1
    100 | 112780 | 23028 | 35878 |       7 |     1 |     1085 |       1
    100 | 112780 | 23028 | 35878 |       8 |     1 |     1197 |       1
    100 | 112780 | 23028 | 35878 |       9 |     2 |      905 |       1
    100 | 112782 | 23028 | 35879 |       0 |     1 |        0 |       0
    100 | 112782 | 23028 | 35879 |       1 |     1 |        0 |       0
    100 | 112782 | 23028 | 35879 |       2 |     1 |        0 |       0
    100 | 112782 | 23028 | 35879 |       3 |     1 |        0 |       0
    100 | 112782 | 23028 | 35879 |       4 |     1 |        0 |       0
    100 | 112782 | 23028 | 35879 |       5 |     1 |        0 |       0
```

```
    100 | 112782 | 23028 | 35879 |       6 |     1 |        0 |       0
    100 | 112782 | 23028 | 35879 |       7 |     1 |        0 |       0
    100 | 112782 | 23028 | 35879 |       8 |     1 |        0 |       0
    100 | 112782 | 23028 | 35879 |       9 |     2 |        0 |       0
(20 rows)
```

## SVCS_CONCURRENCY_SCALING_USAGE

Registra los periodos de uso del escalado de simultaneidad. Cada periodo de uso es una duración consecutiva en la que un clúster de escalado de simultaneidad procesa consultas.

SVCS_CONCURRENCY_SCALING_USAGE Los superusuarios pueden ver esta tabla. El superusuario de la base de datos puede elegir hacerla visible a todos los usuarios.

Columnas de la tabla

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| start_time | TIMESTAMP sin zona horaria | Cuándo comienza el periodo de uso. |
| end_time | TIMESTAMP sin zona horaria | Cuándo termina el periodo de uso. |
| queries | bigint | Número de consultas ejecutadas durante este periodo de uso. |
| usage_in_seconds | numeric (27,0) | Total de segundos en este periodo de uso. |

Consultas de ejemplo

Para ver la duración de uso en segundos para un periodo específico, escriba la siguiente consulta:

```
select * from svcs_concurrency_scaling_usage order by start_time;
```

```
start_time | end_time | queries | usage_in_seconds
---------------------------+---------------------------+---------+-----------------
2019-02-14 18:43:53.01063 | 2019-02-14 19:16:49.781649 | 48 | 1977
```

## SVCS_EXPLAIN

Muestra el plan EXPLAIN de una consulta que se ha enviado para ejecutarse.

> **Note**
>
> Las vistas del sistema con el prefijo SVCS proporcionan información sobre las consultas de los clústeres principal y de escalado de simultaneidad. Las vistas son similares a las tablas con el prefijo STL, con la salvedad de que las tablas STL solo proporcionan información para las consultas ejecutadas en el clúster principal.

SVCS_EXPLAIN es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| userid | integer | ID del usuario que generó la entrada. |
| consulta | integer | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| nodeid | integer | Identificador del nodo del plan, donde se asigna un nodo a uno o más pasos en la ejecución de la consulta. |
| parentid | integer | Identificador del nodo del plan para un nodo principal. Un nodo principal tiene algunos nodos secundarios. Por ejemplo, una combinación de fusión es el nodo principal de los exámenes de las tablas combinadas. |

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| plannode | character (400) | El texto del nodo extraído de la salida de EXPLAIN. Los nodos del plan que hacen referencia a la ejecución en nodos de computación llevan el prefijo **XN** en la salida de EXPLAIN. |
| info | character(400) | Información del calificador y del filtro para el nodo del plan. Por ejemplo, las condiciones de combinación y las restricci ones de la cláusula WHERE están incluidas en esta columna. |

Consultas de ejemplo

Analice la siguiente salida de EXPLAIN para una consulta de combinación de agregación:

```
explain select avg(datediff(day, listtime, saletime)) as avgwait
from sales, listing where sales.listid = listing.listid;
                                QUERY PLAN

------------------------------------------------------------------------------
 XN Aggregate  (cost=6350.30..6350.31 rows=1 width=16)
  ->  XN Hash Join DS_DIST_NONE  (cost=47.08..6340.89 rows=3766 width=16)
        Hash Cond: ("outer".listid = "inner".listid)
        -> XN Seq Scan on listing  (cost=0.00..1924.97 rows=192497 width=12)
        -> XN Hash  (cost=37.66..37.66 rows=3766 width=12)
              -> XN Seq Scan on sales  (cost=0.00..37.66 rows=3766 width=12)
(6 rows)
```

Si ejecuta esta consulta y su ID de consulta es 10, puede utilizar la tabla SVCS_EXPLAIN para ver el mismo tipo de información que devuelve el comando EXPLAIN:

```
select query,nodeid,parentid,substring(plannode from 1 for 30),
substring(info from 1 for 20) from svcs_explain
where query=10 order by 1,2;

query| nodeid |parentid|           substring            |    substring
-----+--------+--------+--------------------------------+------------------
10   |      1 |      0 |XN Aggregate  (cost=6717.61..6  |
10   |      2 |      1 |  -> XN Merge Join DS_DIST_NO| Merge Cond:("outer"
10   |      3 |      2 |       -> XN Seq Scan on lis |
```

```
10   |      4 |      2 |       -> XN Seq Scan on sal |
(4 rows)
```

Analice la siguiente consulta:

```
select event.eventid, sum(pricepaid)
from event, sales
where event.eventid=sales.eventid
group by event.eventid order by 2 desc;

eventid |   sum
--------+----------
    289 | 51846.00
   7895 | 51049.00
   1602 | 50301.00
    851 | 49956.00
   7315 | 49823.00
...
```

Si el ID de esta consulta es 15, la siguiente consulta a la tabla de sistema devolverá los nodos del plan que se realizaron. En este caso, el orden de los nodos está invertido para mostrar el orden rea de la ejecución:

```
select query,nodeid,parentid,substring(plannode from 1 for 56)
from svcs_explain where query=15 order by 1, 2 desc;

query|nodeid|parentid|                          substring
-----+------+--------+-----------------------------------------------------------
15   |    8 |      7 |                                -> XN Seq Scan on eve
15   |    7 |      5 |                          -> XN Hash(cost=87.98..87.9
15   |    6 |      5 |                          -> XN Seq Scan on sales(cos
15   |    5 |      4 |                    -> XN Hash Join DS_DIST_OUTER(cos
15   |    4 |      3 |              -> XN HashAggregate(cost=862286577.07..
15   |    3 |      2 |        -> XN Sort(cost=1000862287175.47..10008622871
15   |    2 |      1 |  -> XN Network(cost=1000862287175.47..1000862287197.
15   |    1 |      0 |XN Merge(cost=1000862287175.47..1000862287197.46 rows=87
(8 rows)
```

La siguiente consulta recupera los ID de consultas de todos los planes de consulta que tienen una función de ventana:

```
select query, trim(plannode) from svcs_explain
```

```
where plannode like '%Window%';

query|                                        btrim
-----+-------------------------------------------------------------------------
26   | -> XN Window(cost=1000985348268.57..1000985351256.98 rows=170 width=33)
27   | -> XN Window(cost=1000985348268.57..1000985351256.98 rows=170 width=33)
(2 rows)
```

## SVCS_PLAN_INFO

Utilice la tabla SVCS_PLAN_INFO para analizar la salida del comando EXPLAIN para una consulta en términos de conjunto de filas. Esta es una manera alternativa de analizar los planes de consulta.

> **Note**
>
> Las vistas del sistema con el prefijo SVCS proporcionan información sobre las consultas de los clústeres principal y de escalado de simultaneidad. Las vistas son similares a las tablas con el prefijo STL, con la salvedad de que las tablas STL solo proporcionan información para las consultas ejecutadas en el clúster principal.

SVCS_PLAN_INFO es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| userid | integer | ID del usuario que generó la entrada. |
| consulta | integer | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| nodeid | integer | Identificador del nodo del plan, donde se asigna un nodo a uno o más pasos en la ejecución de la consulta. |
| segment | integer | Número que identifica el segmento de la consulta. |

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| paso | integer | Número que identifica el paso de la consulta. |
| locus | integer | Ubicación en la que se ejecuta el paso. 0 si es en un nodo de informática y 1 si es en el nodo principal. |
| plannode | integer | Valor enumerado del nodo del plan. Para ver enums de la columna plannode, consulte la siguiente tabla. (La columna PLANNODE en SVCS_EXPLAIN tiene el texto del nodo del plan). |
| startupcost | double precision | El costo relativo estimado de devolver la primera fila de este paso. |
| totalcost | double precision | El costo relativo estimado de ejecutar este paso. |
| rows | bigint | La cantidad estimada de filas que producirá el paso. |
| bytes | bigint | La cantidad estimada de bytes que producirá el paso. |

Consultas de ejemplo

Los siguientes ejemplos comparan los planes de consulta para una consulta simple SELECT, resultado del uso del comando EXPLAIN y de consultar la tabla SVCS_PLAN_INFO.

```
explain select * from category;
QUERY PLAN
------------------------------------------------------------
XN Seq Scan on category (cost=0.00..0.11 rows=11 width=49)
(1 row)

select * from category;
catid | catgroup | catname | catdesc
-------+----------+-----------+--------------------------------------------
1 | Sports | MLB | Major League Baseball
3 | Sports | NFL | National Football League
5 | Sports | MLS | Major League Soccer
...
```

```
select * from svcs_plan_info where query=256;

query | nodeid | segment | step | locus | plannode | startupcost | totalcost
| rows | bytes
-------+--------+---------+------+-------+----------+-------------+-----------+------
+-------
256 | 1 | 0 | 1 | 0 | 104 | 0 | 0.11 | 11 | 539
256 | 1 | 0 | 0 | 0 | 104 | 0 | 0.11 | 11 | 539
(2 rows)
```

En este ejemplo, PLANNODE 104 hace referencia al examen secuencial de la tabla CATEGORY.

```
select distinct eventname from event order by 1;

eventname
----------------------------------------------------------------------------
.38 Special
3 Doors Down
70s Soul Jam
A Bronx Tale
...

explain select distinct eventname from event order by 1;

QUERY PLAN
------------------------------------------------------------------------------------------
XN Merge (cost=1000000000136.38..1000000000137.82 rows=576 width=17)
Merge Key: eventname
-> XN Network (cost=1000000000136.38..1000000000137.82 rows=576
width=17)
Send to leader
-> XN Sort (cost=1000000000136.38..1000000000137.82 rows=576
width=17)
Sort Key: eventname
-> XN Unique (cost=0.00..109.98 rows=576 width=17)
-> XN Seq Scan on event (cost=0.00..87.98 rows=8798
width=17)
(8 rows)

select * from svcs_plan_info where query=240 order by nodeid desc;

query | nodeid | segment | step | locus | plannode | startupcost |
```

```
totalcost | rows | bytes
-------+--------+---------+------+-------+----------+-----------------
+------------------+------+--------
240 | 5 | 0 | 0 | 0 | 104 | 0                | 87.98   | 8798 | 149566
240 | 5 | 0 | 1 | 0 | 104 | 0                | 87.98   | 8798 | 149566
240 | 4 | 0 | 2 | 0 | 117 | 0                | 109.975 | 576  | 9792
240 | 4 | 0 | 3 | 0 | 117 | 0                | 109.975 | 576  | 9792
240 | 4 | 1 | 0 | 0 | 117 | 0                | 109.975 | 576  | 9792
240 | 4 | 1 | 1 | 0 | 117 | 0                | 109.975 | 576  | 9792
240 | 3 | 1 | 2 | 0 | 114 | 1000000000136.38 | 1000000000137.82 | 576 | 9792
240 | 3 | 2 | 0 | 0 | 114 | 1000000000136.38 | 1000000000137.82 | 576 | 9792
240 | 2 | 2 | 1 | 0 | 123 | 1000000000136.38 | 1000000000137.82 | 576 | 9792
240 | 1 | 3 | 0 | 0 | 122 | 1000000000136.38 | 1000000000137.82 | 576 | 9792
(10 rows)
```

## SVCS_QUERY_SUMMARY

Utilice la vista SVCS_QUERY_SUMMARY para encontrar información general acerca de la ejecución de una consulta.

Tenga en cuenta que la información en SVCS_QUERY_SUMMARY proviene de todos los nodos.

> **Note**
>
> La vista SVCS_QUERY_SUMMARY solo tiene información relacionada con las consultas completadas por Amazon Redshift, no otros comandos de utilidades ni de DDL. Para obtener una lista completa e información sobre todas las instrucciones completadas por Amazon Redshift, incluidos los comandos de DDL y de utilidades, puede consultar la vista SVL_STATEMENTTEXT.
>
> Las vistas del sistema con el prefijo SVCS proporcionan información sobre las consultas de los clústeres principal y de escalado de simultaneidad. Las vistas son similares a las vista con el prefijo SVL, con la salvedad de que las vistas SVL solo proporcionan información para las consultas ejecutadas en el clúster principal.

Todos los usuarios pueden ver SVCS_QUERY_SUMMARY. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Para obtener información sobre SVL_QUERY_SUMMARY, consulte SVL_QUERY_SUMMARY.

Columnas de la tabla

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| userid | integer | ID del usuario que generó la entrada. |
| consulta | integer | ID de la consulta. Se puede usar para combinar distintas vistas y tablas del sistema. |
| stm | integer | Secuencia: un conjunto de segmentos simultáneos en una consulta. Una consulta tiene una o más secuencias. |
| seg | integer | Número del segmento. Una consulta consta de distintos segmentos y cada segmento consta de uno o más pasos. Los segmentos de una consulta se ejecutan en paralelo. Cada segmento se ejecuta en un solo proceso. |
| paso | integer | Paso de la consulta que se ejecutó. |
| maxtime | bigint | Cantidad máxima de tiempo en que se ejecuta el paso (en microsegundos). |
| avgtime | bigint | Tiempo promedio en que se ejecuta el paso (en microsegundos). |
| rows | bigint | Cantidad de datos involucrados en el paso de la consulta. |
| bytes | bigint | Cantidad de bytes de datos involucrados en el paso de la consulta. |
| rate_row | double precision | Proporción de ejecución de la consulta por fila. |
| rate_byte | double precision | Proporción de ejecución de la consulta por byte. |

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| etiqueta | texto | Etiqueta del paso, que se compone del nombre de un paso de la consulta y, si corresponde, el ID y el nombre de la tabla (por ejemplo, scan tbl=100448 name =user). Los ID de tabla de tres dígitos suelen referirse a exámenes de tablas transitorias. Cuando vea `tbl=0`, por lo general se refiere a un examen de un valor constante. |
| is_diskbased | character (1) | Indica si este paso de la consulta se realizó como una operación basada en el disco en algún nodo del clúster: true (**t**) (verdadero) o false (**f**) (falso). Solo determinados pasos, como hash, ordenar y agregar pasos, pueden ir al disco. Muchos tipos de pasos se ejecutan siempre en la memoria. |
| workmem | bigint | Cantidad de memoria funcional (en bytes) asignada a este paso de la consulta. |
| is_rrscan | character (1) | Si es true (**t**), indica que se utilizó un examen de rango restringido en el paso. El valor predeterminado es false (**f**). |
| is_delayed_scan | character (1) | Si es true (**t**), indica que en el paso se utilizó un examen diferido. El valor predeterminado es false (**f**). |
| rows_pre_filter | bigint | Para exámenes de tablas permanentes, la cantidad total de filas emitidas antes de filtrar las filas marcadas para eliminación (filas fantasma). |

Consultas de ejemplo

Vista de la información de procesamiento del paso de una consulta

En la siguiente consulta, se muestra la información básica de procesamiento para cada paso de la consulta 87:

```
select query, stm, seg, step, rows, bytes
from svcs_query_summary
where query = 87
```

```
order by query, seg, step;
```

Esta consulta recupera la información de procesamiento acerca de la consulta 87, tal como se muestra en el siguiente ejemplo de salida:

```
 query | stm | seg | step |  rows  |  bytes
-------+-----+-----+------+--------+---------
87     |   0 |   0 |    0 |     90 |    1890
87     |   0 |   0 |    2 |     90 |     360
87     |   0 |   1 |    0 |     90 |     360
87     |   0 |   1 |    2 |     90 |    1440
87     |   1 |   2 |    0 | 210494 | 4209880
87     |   1 |   2 |    3 |  89500 |       0
87     |   1 |   2 |    6 |      4 |      96
87     |   2 |   3 |    0 |      4 |      96
87     |   2 |   3 |    1 |      4 |      96
87     |   2 |   4 |    0 |      4 |      96
87     |   2 |   4 |    1 |      1 |      24
87     |   3 |   5 |    0 |      1 |      24
87     |   3 |   5 |    4 |      0 |       0
(13 rows)
```

Determinar qué pasos de la consulta se vertieron en el disco

En la siguiente consulta, se muestra si alguno de los pasos de la consulta con el ID de consulta 1025 (ver la vista SVL_QLOG para saber cómo obtener el ID de una consulta) se vertieron al disco o si la consulta se ejecutó en su totalidad en la memoria:

```
select query, step, rows, workmem, label, is_diskbased
from svcs_query_summary
where query = 1025
order by workmem desc;
```

Esta consulta devuelve el siguiente ejemplo de salida:

```
query| step|  rows  |  workmem   |  label        | is_diskbased
-----+-----+--------+------------+---------------+--------------
1025 |  0  |16000000|  141557760 |scan tbl=9     | f
1025 |  2  |16000000|  135266304 |hash tbl=142   | t
1025 |  0  |16000000|  128974848 |scan tbl=116536| f
1025 |  2  |16000000|  122683392 |dist           | f
```

```
(4 rows)
```

Al examinar los valores para IS_DISKBASED, puede ver qué pasos de la consulta pasaron al disco. Para la consulta 1025, el paso hash se ejecutó en el disco. Los pasos que se pueden ejecutar en el disco incluyen los pasos hash, aggr y sort. Para ver únicamente los pasos de la consulta basados en el disco, agregue la cláusula **`and is_diskbased = 't'`** a la instrucción SQL del ejemplo anterior.

## SVCS_S3LIST

Utilice la vista SVCS_S3LIST para obtener detalles acerca de las consultas de Amazon Redshift Spectrum en el nivel de segmento. Un segmento puede realizar un análisis de tablas externas. Esta vista se obtiene de la vista del sistema SVL_S3LIST, pero no muestra el sector de las consultas ejecutadas en un clúster de escalado de simultaneidad.

> **Note**
>
> Las vistas del sistema con el prefijo SVCS proporcionan información sobre las consultas de los clústeres principal y de escalado de simultaneidad. Las vistas son similares a las vista con el prefijo SVL, con la salvedad de que las vistas SVL solo proporcionan información para las consultas ejecutadas en el clúster principal.

Todos los usuarios pueden ver SVCS_S3LIST. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Para obtener más información sobre SVL_S3LIST, consulte SVL_S3LIST.

Columnas de la tabla

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| consulta | integer | El ID de la consulta. |
| segment | integer | El número del segmento. Una consulta consta de varios segmentos. |
| node | integer | El número de nodo. |

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| eventtime | Marca de tiempo | La hora en UTC a la que se registra el evento. |
| bucket | char (256) | El nombre del bucket de Amazon S3. |
| prefix | char (256) | El prefijo de la ubicación del bucket de Amazon S3. |
| recursive | char(1) | Si se realiza un análisis recursivo de las subcarpetas. |
| retrieved_files | integer | El número de archivos mostrados. |
| max_file_size | bigint | El tamaño de archivo máximo entre los archivos mostrados. |
| avg_file_size | double precision | El tamaño de archivo medio entre los archivos mostrados. |
| generated_splits | integer | El número de divisiones de archivo. |
| avg_split_length | double precision | La longitud media de las divisiones de archivos en bytes. |
| duration | bigint | La duración del listado de archivos, en microsegundos. |

Consulta de ejemplo

En el siguiente ejemplo, se consulta SVCS_S3LIST para obtener la última consulta realizada.

```
select *
from svcs_s3list
where query = pg_last_query_id()
order by query,segment;
```

# SVCS_S3LOG

Utilice la vista SVCS_S3LOG para obtener detalles acerca de la resolución de problemas de las consultas de Redshift Spectrum en el nivel de segmento. Un segmento puede realizar un análisis de tablas externas. Esta vista se obtiene de la vista del sistema SVL_S3LOG, pero no muestra el sector de las consultas ejecutadas en un clúster de escalado de simultaneidad.

> **Note**
>
> Las vistas del sistema con el prefijo SVCS proporcionan información sobre las consultas de los clústeres principal y de escalado de simultaneidad. Las vistas son similares a las vista con el prefijo SVL, con la salvedad de que las vistas SVL solo proporcionan información para las consultas ejecutadas en el clúster principal.

Todos los usuarios pueden ver SVCS_S3LOG. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Para obtener más información sobre SVL_S3LOG, consulte SVL_S3LOG.

Columnas de la tabla

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| pid | integer | El ID del proceso. |
| consulta | integer | El ID de la consulta. |
| segment | integer | El número del segmento. Una consulta consta de distintos segmentos y cada segmento consta de uno o más pasos. |
| paso | integer | El paso de la consulta que se ejecutó. |
| node | integer | El número de nodo. |
| eventtime | Marca de tiempo | La hora en UTC a la que se registra el evento. |
| message | char(512) | Mensaje para la entrada del registro. |

Consulta de ejemplo

En el siguiente ejemplo, se consulta SVCS_S3LOG para obtener la última consulta ejecutada.

```
select *
from svcs_s3log
where query = pg_last_query_id()
order by query,segment;
```

## SVCS_S3PARTITION_SUMMARY

Utilice la vista SVCS_S3PARTITION_SUMMARY para obtener un resumen de la partición de consultas de Redshift Spectrum que se procesa en el nivel de segmento. Un segmento puede realizar un análisis de tablas externas.

> **Note**
>
> Las vistas del sistema con el prefijo SVCS proporcionan información sobre las consultas de los clústeres principal y de escalado de simultaneidad. Las vistas son similares a las vista con el prefijo SVL, con la salvedad de que las vistas SVL solo proporcionan información para las consultas ejecutadas en el clúster principal.

Todos los usuarios pueden ver SVCS_S3PARTITION_SUMMARY. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Para obtener más información sobre SVL_S3PARTITION, consulte SVL_S3PARTITION.

Columnas de la tabla

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| consulta | integer | El ID de la consulta. Puede usar este valor para combinar distintas vistas y tablas del sistema. |
| segment | integer | El número del segmento. Una consulta consta de varios segmentos. |

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| assignment | char(1) | El tipo de asignación de partición entre nodos. |
| min_starttime | Marca de tiempo | Hora en UTC a la que se inició el procesamiento de la partición. |
| max_endtime | Marca de tiempo | Hora en UTC a la que finalizó el procesamiento de la partición. |
| min_duration | bigint | Tiempo mínimo de procesamiento de la partición usado por un nodo para esta consulta (en microsegundos). |
| max_duration | bigint | Tiempo máximo de procesamiento de la partición usado por un nodo para esta consulta (en microsegundos). |
| avg_duration | bigint | Tiempo medio de procesamiento de la partición usado por un nodo para esta consulta (en microsegundos). |
| total_partitions | integer | Número total de particiones en una tabla externa. |
| qualified_partitions | integer | Número total de particiones aptas. |
| min_assigned_partitions | integer | Número mínimo de particiones asignadas en un nodo. |
| max_assigned_partitions | integer | Número máximo de particiones asignadas en un nodo. |
| avg_assigned_partitions | bigint | Número medio de particiones asignadas en un nodo. |

Consulta de ejemplo

En el siguiente ejemplo, se obtienen los detalles del análisis de partición de la última consulta realizada.

```
select query, segment, assignment, min_starttime, max_endtime, min_duration,
 avg_duration
from svcs_s3partition_summary
where query = pg_last_query_id()
order by query,segment;
```

## SVCS_S3QUERY_SUMMARY

Utilice la vista SVCS_S3QUERY_SUMMARY para obtener un resumen de todas las consultas de Redshift Spectrum (consultas de S3) que se ejecutaron en el sistema. Un segmento puede realizar un análisis de tablas externas.

> **Note**
>
> Las vistas del sistema con el prefijo SVCS proporcionan información sobre las consultas de los clústeres principal y de escalado de simultaneidad. Las vistas son similares a las vista con el prefijo SVL, con la salvedad de que las vistas SVL solo proporcionan información para las consultas ejecutadas en el clúster principal.

Todos los usuarios pueden ver SVCS_S3QUERY_SUMMARY. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Para obtener más información sobre SVL_S3QUERY, consulte SVL_S3QUERY.

Columnas de la tabla

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| userid | integer | ID del usuario que generó la entrada dada. |
| consulta | integer | El ID de la consulta. Puede usar este valor para combinar distintas vistas y tablas del sistema. |

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| xid | bigint | El ID de la transacción. |
| pid | integer | El ID del proceso. |
| segment | integer | El número del segmento. Una consulta consta de distintos segmentos y cada segmento consta de uno o más pasos. |
| paso | integer | El paso de la consulta que se ejecutó. |
| starttime | Marca de tiempo | La hora en UTC a la que comenzó a ejecutarse la consulta de Redshift Spectrum en este segmento. Un segmento puede tener un análisis de tablas externas. |
| endtime | Marca de tiempo | La hora en UTC a la que terminó de ejecutarse la consulta de Redshift Spectrum en este segmento. Un segmento puede tener un análisis de tablas externas. |
| elapsed | integer | El tiempo que tardó la consulta de Redshift Spectrum en ejecutarse en este segmento (en microsegundos). |
| aborted | integer | Si el sistema detuvo la consulta o el usuario la canceló, esta columna tendrá el valor **1**. Si la consulta se ejecutó hasta su finalización, esta columna tendrá el valor **0**. |
| external_table_name | char (136) | El formato interno del nombre del nombre externo de la tabla para el examen de la tabla externa. |
| file_format | character(16) | El formato de archivo de los datos de la tabla externa. |
| is_partitioned | char(1) | Si es true (**t**), el valor de esta columna indica que la tabla externa está particionada. |
| is_rrscan | char (1) | Si es true (**t**), el valor de esta columna indica que se aplicó un examen de rango restringido. |

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| is_nested | varchar(1) | Si es true (**t**), el valor de esta columna indica que se obtuvo acceso al tipo de datos de la columna anidada. |
| s3_scanned_rows | bigint | La cantidad de filas examinadas desde Amazon S3 y enviadas a la capa de Redshift Spectrum. |
| s3_scanned_bytes | bigint | La cantidad de bytes examinados desde Amazon S3 y enviados a la capa de Redshift Spectrum, según los datos comprimidos. |
| s3query_returned_rows | bigint | La cantidad de filas devueltas desde la capa Redshift Spectrum hasta el clúster. |
| s3query_returned_bytes | bigint | La cantidad de bytes devueltos desde la capa Redshift Spectrum hasta el clúster. Una gran cantidad de datos devueltos a Amazon Redshift podría afectar el rendimiento del sistema. |
| files | integer | La cantidad de filas que se procesaron para esta consulta de Redshift Spectrum. Una pequeña cantidad de archivos limita los beneficios del procesamiento en paralelo. |
| files_max | integer | Cantidad total de archivos procesados en un sector. |
| files_avg | integer | Cantidad promedio de archivos procesados en un sector. |
| splits | bigint | Número de divisiones procesadas para este segmento. Número de divisiones procesadas en este sector. En el caso de los archivos de datos divisibles de gran tamaño (por ejemplo, archivos que superan los 512 MB), Redshift Spectrum intenta dividir los archivos en varias solicitudes de S3 para procesarlos en paralelo. |
| splits_max | integer | Número máximo de divisiones procesadas en este sector. |

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| splits_avg | bigint | Número medio de divisiones procesadas en este sector. |
| total_split_size | bigint | Tamaño total de todas las divisiones procesadas. |
| max_split_size | bigint | Tamaño máximo en bytes de las divisiones procesadas. |
| avg_split_size | bigint | Tamaño medio en bytes de las divisiones procesadas. |
| total_retries | bigint | La cantidad total de reintentos de la consulta de Redshift Spectrum en este segmento. |
| max_retries | integer | Cantidad máxima de reintentos para un archivo individual procesado. |
| max_request_duration | bigint | La duración máxima de una solicitud de un archivo individual (en microsegundos). Si las consultas tienen demasiado tiempo de ejecución, esto podría deberse a un cuello de botella. |
| avg_request_duration | bigint | La duración media de las solicitudes de archivo (en microsegundos). |
| max_request_parallelism | integer | La cantidad máxima de solicitudes en paralelo en un sector para esta consulta de Redshift Spectrum. |
| avg_request_parallelism | double precision | La cantidad promedio de solicitudes en paralelo en un sector para esta consulta de Redshift Spectrum. |
| total_slowdown_count | bigint | La cantidad total de solicitudes de Amazon S3 con un error de procesamiento lento que se produjeron durante el análisis de la tabla externa. |

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| max_slowdown_count | integer | La cantidad máxima de solicitudes de Amazon S3 con un error de procesamiento lento que se produjeron durante el análisis de la tabla externa en un sector. |

Consulta de ejemplo

En el siguiente ejemplo, se obtienen los detalles del paso de examen para la última consulta ejecutada.

```
select query, segment, elapsed, s3_scanned_rows, s3_scanned_bytes,
 s3query_returned_rows, s3query_returned_bytes, files
from svcs_s3query_summary
where query = pg_last_query_id()
order by query,segment;
```

```
query | segment | elapsed | s3_scanned_rows | s3_scanned_bytes | s3query_returned_rows
 | s3query_returned_bytes | files
------+---------+---------+-----------------+------------------+----------------------
+------------------------+------
 4587 |       2 |   67811 |               0 |                0 |                    0
 |                      0 |     0
 4587 |       2 |  591568 |          172462 |         11260097 |                 8513
 |                 170260 |     1
 4587 |       2 |  216849 |               0 |                0 |                    0
 |                      0 |     0
 4587 |       2 |  216671 |               0 |                0 |                    0
 |                      0 |     0
```

## SVCS_STREAM_SEGS

Muestra la relación entre las secuencias y los segmentos simultáneos.

> **Note**
>
> Las vistas del sistema con el prefijo SVCS proporcionan información sobre las consultas de los clústeres principal y de escalado de simultaneidad. Las vistas son similares a las tablas

con el prefijo STL, con la salvedad de que las tablas STL solo proporcionan información para las consultas ejecutadas en el clúster principal.

SVCS_STREAM_SEGS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| userid | integer | ID del usuario que generó la entrada. |
| consulta | integer | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| secuencia | integer | El conjunto de segmentos simultáneos de una consulta. |
| segment | integer | Número que identifica el segmento de la consulta. |

Consultas de ejemplo

Para ver la relación entre las secuencias y los segmentos simultáneos de la consulta más reciente, escriba la siguiente consulta:

```
select *
from svcs_stream_segs
where query = pg_last_query_id();

 query | stream | segment
-------+--------+---------
    10 |      1 |       2
    10 |      0 |       0
    10 |      2 |       4
    10 |      1 |       3
    10 |      0 |       1
(5 rows)
```

# SVCS_UNLOAD_LOG

Utilice SVCS_UNLOAD_LOG para obtener los detalles de las operaciones UNLOAD.

SVCS_UNLOAD_LOG registra una fila por cada archivo creado por una instrucción UNLOAD. Por ejemplo, si una instrucción UNLOAD crea 12 archivos, SVCS_UNLOAD_LOG tendrá 12 filas correspondientes. Esta vista se obtiene de la tabla del sistema STL_UNLOAD_LOG, pero no muestra el sector de las consultas ejecutadas en un clúster de escalado de simultaneidad.

> **Note**
>
> Las vistas del sistema con el prefijo SVCS proporcionan información sobre las consultas de los clústeres principal y de escalado de simultaneidad. Las vistas son similares a las tablas con el prefijo STL, con la salvedad de que las tablas STL solo proporcionan información para las consultas ejecutadas en el clúster principal.

SVCS_UNLOAD_LOG es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| userid | integer | ID del usuario que generó la entrada. |
| consulta | integer | El ID de la consulta. |
| pid | integer | ID de proceso asociado a la instrucción de la consulta. |
| ruta | character (1280) | La ruta competa del objeto de Amazon S3 para el archivo. |
| start_time | Marca de tiempo | La hora de inicio de la operación UNLOAD. |
| end_time | Marca de tiempo | La hora de finalización de la operación de UNLOAD. |
| line_count | bigint | Cantidad de líneas (filas) descargadas en el archivo. |

| Nombre de la columna | Tipo de datos | Descripción |
|---|---|---|
| transfer_size | bigint | Número de bytes transferidos. |
| file_format | character(10) | El formato de archivo no cargado. |

Consulta de ejemplo

Para obtener una lista de los archivos que se escribieron en Amazon S3 mediante un comando UNLOAD, puede llamar a una operación de lista de Amazon S3 una vez que se complete el comando UNLOAD. Sin embargo, en función de la rapidez con la que se emita la llamada, la lista podría estar incompleta, ya que las operaciones de lista de Amazon S3 tienen consistencia final. Para obtener una lista oficial completa de inmediato, consulte SVCS_UNLOAD_LOG.

La siguiente consulta devuelve el nombre de la ruta de los archivos que se crearon a través de un comando UNLOAD para la última consulta completada:

```
select query, substring(path,0,40) as path
from svcs_unload_log
where query = pg_last_query_id()
order by path;
```

Este comando devuelve la siguiente salida de ejemplo:

```
 query |             path
 ------+-----------------------------------------
  2320 | s3://amzn-s3-demo-bucket/venue0000_part_00
  2320 | s3://amzn-s3-demo-bucket/venue0001_part_00
  2320 | s3://amzn-s3-demo-bucket/venue0002_part_00
  2320 | s3://amzn-s3-demo-bucket/venue0003_part_00
(4 rows)
```

## Vistas de SVL para el clúster principal

Las vistas de SVL son vistas del sistema en Amazon Redshift que contienen referencias a tablas y registros STL donde puede obtener información más detallada.

Estas vistas proporcionan un acceso más rápido y fácil a los datos más consultados que se encuentran en esas tablas.

> **Note**
>
> La vista SVL_QUERY_SUMMARY solo contiene información relacionada con las consultas ejecutadas por Amazon Redshift, no otros comandos de utilidades ni de DDL. Para obtener una lista completa e información acerca de todas las instrucciones ejecutadas por Amazon Redshift, incluidos los comandos de DDL y de utilidades, puede consultar la vista SVL_STATEMENTTEXT.

Temas


- SVL_AUTO_WORKER_ACTION
- SVL_COMPILE
- SVL_DATASHARE_CHANGE_LOG
- SVL_DATASHARE_CROSS_REGION_USAGE
- SVL_DATASHARE_USAGE_CONSUMER
- SVL_DATASHARE_USAGE_PRODUCER
- SVL_FEDERATED_QUERY
- SVL_MULTI_STATEMENT_VIOLATIONS
- SVL_MV_REFRESH_STATUS
- SVL_QERROR
- SVL_QLOG
- SVL_QUERY_METRICS
- SVL_QUERY_METRICS_SUMMARY
- SVL_QUERY_QUEUE_INFO
- SVL_QUERY_REPORT
- SVL_QUERY_SUMMARY
- SVL_RESTORE_ALTER_TABLE_PROGRESS
- SVL_S3LIST
- SVL_S3LOG

- SVL_S3PARTITION
- SVL_S3PARTITION_SUMMARY
- SVL_S3QUERY
- SVL_S3QUERY_SUMMARY
- SVL_S3RETRIES
- SVL_SPATIAL_SIMPLIFY
- SVL_SPECTRUM_SCAN_ERROR
- SVL_STATEMENTTEXT
- SVL_STORED_PROC_CALL
- SVL_STORED_PROC_MESSAGES
- SVL_TERMINATE
- SVL_UDF_LOG
- SVL_USER_INFO
- SVL_VACUUM_PERCENTAGE

## SVL_AUTO_WORKER_ACTION

Registra las acciones automatizadas realizadas por Amazon Redshift en tablas definidas para la optimización automática.

SVL_AUTO_WORKER_ACTION es visible solo para los superusuarios. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| table_id | entero | El identificador de tabla. |
| type | character (32) | El tipo de recomendación. Los valores posibles son distkey y sortkey. |
| status | character (128) | El estado de realización de la recomendación. Los valores posibles son Start, Complete, Omited, Abort, Checkpoint y Failed. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| eventtime | marca de tiempo | La marca temporal de la columna de `status`. |
| sequence | entero | El número de secuencia de un valor `previous_state` truncado. Cuando un `previous_state` único tiene más de 200 caracteres, se registran filas adicionales para ese valor. La secuencia es la siguiente: 0 es la primera fila, 1 es la segunda y, así, sucesivamente. |
| previous_state | character (200) | El estilo de distribución anterior y las claves de ordenación de la tabla antes de aplicar la recomendación. El valor se trunca en incrementos de 200 caracteres. |

Algunos ejemplos de valores de la columna `status` son los siguientes:

- Skipped:Table not found.
- Skipped:Recommendation is empty.
- Skipped:Apply sortkey recommendation is disabled.
- Skipped:Retry exceeds the maximum limit for a table.
- Skipped:Table column has changed.
- Abort:This table is not AUTO.
- Abort:This table has been recently converted.
- Abort:This table exceeds table size threshold.
- Abort:This table is already the recommended style.
- Checkpoint: progress *21.9963*%.

Consultas de ejemplo

En el siguiente ejemplo, las filas del resultado muestran las acciones realizadas por Amazon Redshift.

```
select table_id, type, status, eventtime, sequence, previous_state
```

```
from SVL_AUTO_WORKER_ACTION;
```

```
 table_id |  type   |                        status                        |
 eventtime          | sequence | previous_state
----------+---------+------------------------------------------------------
+----------------------------+----------+----------------
   118082 | sortkey | Start                                                | 2020-08-22
 19:42:20.727049 | 0        |
   118078 | sortkey | Start                                                | 2020-08-22
 19:43:54.728819 | 0        |
   118082 | sortkey | Start                                                | 2020-08-22
 19:42:52.690264 | 0        |
   118072 | sortkey | Start                                                | 2020-08-22
 19:44:14.793572 | 0        |
   118082 | sortkey | Failed                                               | 2020-08-22
 19:42:20.728917 | 0        |
   118078 | sortkey | Complete                                             | 2020-08-22
 19:43:54.792705 | 0        | SORTKEY: None;
   118086 | sortkey | Complete                                             | 2020-08-22
 19:42:00.72635  | 0        | SORTKEY: None;
   118082 | sortkey | Complete                                             | 2020-08-22
 19:43:34.728144 | 0        | SORTKEY: None;
   118072 | sortkey | Skipped:Retry exceeds the maximum limit for a table. | 2020-08-22
 19:44:46.706155 | 0        |
   118086 | sortkey | Start                                                | 2020-08-22
 19:42:00.685255 | 0        |
   118082 | sortkey | Start                                                | 2020-08-22
 19:43:34.69531  | 0        |
   118072 | sortkey | Start                                                | 2020-08-22
 19:44:46.703331 | 0        |
   118082 | sortkey | Checkpoint: progress 14.755079%                      | 2020-08-22
 19:42:52.692828 | 0        |
   118072 | sortkey | Failed                                               | 2020-08-22
 19:44:14.796071 | 0        |
   116723 | sortkey | Abort:This table is not AUTO.                        | 2020-10-28
 05:12:58.479233 | 0        |
   110203 | distkey | Abort:This table is not AUTO.                        | 2020-10-28
 05:45:54.67259  | 0        |
```

## SVL_COMPILE

Registra el tiempo de compilación y la ubicación de cada segmento de una determinada consulta.

SVL_COMPILE es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> SVL_COMPILE solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_QUERY_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Para obtener más información sobre SVCS_COMPILE, consulte SVCS_COMPILE.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| xid | bigint | ID de la transacción asociado a la instrucción. |
| pid | entero | ID del proceso asociado a la instrucción. |
| consulta | entero | ID de la consulta. Se puede usar para combinar distintas vistas y tablas del sistema. |
| segment | entero | El segmento de la consulta que se compilará. |
| locus | entero | Ubicación en la que se ejecuta el segmento. **1** si es en un nodo de informática y **2** si es en el nodo principal. |
| starttime | marca de tiempo | Hora en UTC en que comenzó la compilación. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| endtime | marca de tiempo | Hora en UTC en que finalizó la compilación. |
| compile | entero | **0** si la compilación se reutilizó, **1** si se compiló el segmento. |

Consultas de ejemplo

En este ejemplo, las consultas 35878 y 35879 ejecutaron la misma instrucción SQL. La columna de compilación para la consulta 35878 muestra 1 para cuatro segmentos de consulta, lo que indica que esos segmentos se compilaron. La consulta 35879 muestra `0` en la columna de compilación para cada segmento, lo que indica que esos segmentos no se volvieron a compilar.

```
select userid, xid,  pid, query, segment, locus,
datediff(ms, starttime, endtime) as duration, compile
from svl_compile
where query = 35878 or query = 35879
order by query, segment;

 userid |  xid   |  pid  | query | segment | locus | duration | compile
--------+--------+-------+-------+---------+-------+----------+---------
    100 | 112780 | 23028 | 35878 |       0 |     1 |        0 |       0
    100 | 112780 | 23028 | 35878 |       1 |     1 |        0 |       0
    100 | 112780 | 23028 | 35878 |       2 |     1 |        0 |       0
    100 | 112780 | 23028 | 35878 |       3 |     1 |        0 |       0
    100 | 112780 | 23028 | 35878 |       4 |     1 |        0 |       0
    100 | 112780 | 23028 | 35878 |       5 |     1 |        0 |       0
    100 | 112780 | 23028 | 35878 |       6 |     1 |     1380 |       1
    100 | 112780 | 23028 | 35878 |       7 |     1 |     1085 |       1
    100 | 112780 | 23028 | 35878 |       8 |     1 |     1197 |       1
    100 | 112780 | 23028 | 35878 |       9 |     2 |      905 |       1
    100 | 112782 | 23028 | 35879 |       0 |     1 |        0 |       0
    100 | 112782 | 23028 | 35879 |       1 |     1 |        0 |       0
    100 | 112782 | 23028 | 35879 |       2 |     1 |        0 |       0
    100 | 112782 | 23028 | 35879 |       3 |     1 |        0 |       0
    100 | 112782 | 23028 | 35879 |       4 |     1 |        0 |       0
    100 | 112782 | 23028 | 35879 |       5 |     1 |        0 |       0
    100 | 112782 | 23028 | 35879 |       6 |     1 |        0 |       0
```

```
    100 | 112782 | 23028 | 35879 |       7 |     1 |        0 |       0
    100 | 112782 | 23028 | 35879 |       8 |     1 |        0 |       0
    100 | 112782 | 23028 | 35879 |       9 |     2 |        0 |       0
(20 rows)
```

## SVL_DATASHARE_CHANGE_LOG

Registra la vista consolidada para realizar el seguimiento de los cambios en los datashares (recursos para compartir datos) tanto en los clústeres productores como en los consumidores.

SVL_DATASHARE_CHANGE_LOG es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_DATASHARE_CHANGE_LOG. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | El ID del grupo de usuarios que realizan la acción. |
| nombre de usuario | varchar(128) | El nombre del usuario que realiza la acción. |
| pid | entero | El ID del proceso. |
| xid | bigint | El ID de la transacción. |
| share_id | entero | El ID del datashare afectado. |
| share_name | varchar(128) | El nombre del datashare. |
| source_database_id | entero | El ID de la base de datos a la que pertenece el datashare. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| source_database_name | varchar(128) | El nombre de la base de datos a la que pertenece el datashare. |
| consumer_database_id | entero | El ID de la base de datos importada desde el datashare. |
| consumer_database_name | varchar(128) | El nombre de la base de datos importada desde el datashare. |
| arn | varchar(192) | El ARN del recurso que respalda la base de datos importada. |
| recordtime | marca de tiempo | La marca temporal de la acción. |
| acción | varchar(128) | La acción que se está ejecutando. Los valores posibles son CREATE DATASHARE, DROP DATASHARE, GRANT ALTER, REVOKE ALTER, GRANT SHARE, REVOKE SHARE, ALTER ADD, ALTER REMOVE, ALTER SET, GRANT USAGE, REVOKE USAGE, CREATE DATABASE, GRANT o REVOKE USAGE en una base de datos compartida, DROP SHARED DATABASE, ALTER SHARED DATABASE. |
| status | entero | El estado de la acción. Los valores posibles son SUCCESS y ERROR-ERROR CODE. |
| share_object_type | varchar (64) | El tipo de objeto de base de datos que se agregó al datashare o se eliminó de él. Los valores posibles son esquemas, tablas, columnas, funciones y vistas. Este es un campo para el clúster productor. |
| share_object_id | entero | El ID del objeto de base de datos que se agregó al datashare o se eliminó de él. Este es un campo para el clúster productor. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| share_object_name | varchar(128) | El nombre del objeto de base de datos que se agregó al datashare o se eliminó de él. Este es un campo para el clúster productor. |
| target_user_type | varchar(16) | El tipo de usuarios o grupos a los que se concedió un privilegio. Este es un campo tanto para el clúster productor como para el consumidor. |
| target_userid | entero | Los ID de usuarios o grupos a los que se concedió un privilegio. Este es un campo tanto para el clúster productor como para el consumidor. |
| target_username | varchar(128) | El nombre de los usuarios o grupos a los que se concedió un privilegio. Este es un campo tanto para el clúster productor como para el consumidor. |
| consumer_account | varchar(16) | El ID de cuenta del consumidor de datos. Este es un campo para el clúster productor. |
| consumer_namespace | varchar(64) | El espacio de nombres de la cuenta consumidora de datos. Este es un campo para el clúster productor. |
| producer_account | varchar(16) | El ID de la cuenta productora a la que pertenece el datashare. Este es un campo para el clúster consumidor. |
| producer_namespace | varchar(64) | El espacio de nombres de la cuenta de producto a la que pertenece el datashare. Este es un campo para el clúster consumidor. |
| attribute_name | varchar(64) | El nombre de un atributo de datashare o base de datos compartida. |
| attribute_value | varchar(128) | El valor de un atributo de datashare o base de datos compartida. |
| mensaje | varchar(512) | El mensaje de error cuando se produce un error en una acción. |

Consultas de ejemplo

En el siguiente ejemplo, se muestra una vista SVL_DATASHARE_CHANGE_LOG.

```
SELECT DISTINCT action
FROM svl_datashare_change_log
WHERE share_object_name LIKE 'tickit%';

         action
 ----------------------
  "ALTER DATASHARE ADD"
```

## SVL_DATASHARE_CROSS_REGION_USAGE

Utilice la vista SVL_DATASHARE_CROSS_REGION_USAGE para obtener un resumen del uso de datos transferidos entre regiones debido a la consulta de uso compartido de datos entre regiones. SVL_DATASHARE_CROSS_REGION_USAGE agrega detalles en el nivel de segmento.

SVL_DATASHARE_CROSS_REGION_USAGE es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_DATASHARE_CROSS_REGION_USAGE. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| consulta | entero | El Id. de la consulta. Utilice este valor para unir otras tablas y vistas del sistema. |
| segment | bigint | Número del segmento. Una consulta consta de distintos segmentos y cada segmento consta de uno o más pasos. |
| start_time | hora | Es la hora en UTC en que se ha iniciado la transferencia de datos. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| end_time | hora | Es la hora en UTC en que ha finalizado la transferencia de datos. |
| transferred_data | bigint | Número de bytes de datos transferidos de una región productora a una región consumidora. |
| source_region | char(25) | Región productora desde la que la consulta transfirió los datos. |
| recordtime | marca de tiempo | La hora en que se registra la acción. |

Consultas de ejemplo

En el siguiente ejemplo, se muestra una vista SVL_DATASHARE_CROSS_REGION_USAGE.

```
SELECT query, segment, transferred_data, source_region
from svl_datashare_cross_region_usage
where query = pg_last_query_id()
order by query,segment;

  query | segment | transferred_data | source_region
--------+---------+------------------+---------------
 200048 |       2 |          4194304 |    us-west-1
 200048 |       2 |          4194304 |    us-east-2
```

## SVL_DATASHARE_USAGE_CONSUMER

Registra la actividad y el uso de los datashares. Esta vista solo es relevante en el clúster consumidor.

SVL_DATASHARE_USAGE_CONSUMER es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_DATASHARE_USAGE_CONSUMER. Los datos de la vista de monitoreo SYS están

formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | El ID del usuario que emite la solicitud. |
| pid | entero | El ID del proceso principal que ejecuta la consulta. |
| xid | bigint | El contexto de la transacción actual. |
| request_id | varchar(50) | El ID único de la llamada a la API solicitada. |
| request_type | varchar(25) | El tipo de la solicitud realizada al clúster productor. |
| transaction_uid | varchar(50) | El ID único de la transacción. |
| recordtime | marca de tiempo | La hora en que se registra la acción. |
| status | entero | El estado de la llamada a la API solicitada. |
| error | varchar(512) | El mensaje de un error. |

Consultas de ejemplo

En el siguiente ejemplo, se muestra una vista SVL_DATASHARE_USAGE_CONSUMER.

```
SELECT request_type, status, trim(error) AS error
FROM svl_datashare_usage_consumer

  request_type  | status | error
```

```
----------------+--------+--------
 "GET RELATION" |   0    |
```

## SVL_DATASHARE_USAGE_PRODUCER

Registra la actividad y el uso de los datashares. Esta vista solo es relevante en el clúster productor.

SVL_DATASHARE_USAGE_PRODUCER es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_DATASHARE_USAGE_PRODUCER. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| share_id | entero | El ID de objeto (OID) del datashare. |
| share_name | varchar(128) | El nombre del datashare. |
| request_id | varchar(50) | El ID único de la llamada a la API solicitada. |
| request_type | varchar(25) | El tipo de la solicitud realizada al clúster productor. |
| object_type | varchar (64) | El tipo de objeto que se comparte desde el datashare. Los valores posibles son esquemas, tablas, columnas, funciones y vistas. |
| object_oid | entero | El ID del objeto que se comparte desde el datashare. |
| object_name | varchar(128) | El nombre del objeto que se comparte desde el datashare. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| consumer_account | varchar (16) | La cuenta de la cuenta consumidora con la que se comparte el datashare. |
| consumer_namespace | varchar (64) | El espacio de nombres de la cuenta consumidora con la que se comparte el datashare. |
| consumer_transaction_uid | varchar(50) | El ID de transacción único de la instrucción en el clúster consumidor. |
| recordtime | marca de tiempo | La hora en que se registra la acción. |
| status | entero | El estado del datashare. |
| error | varchar(512) | El mensaje de un error. |
| consumer_region | char(64) | Es la región en la que se encuentra el clúster del consumidor. |

Consultas de ejemplo

En el siguiente ejemplo, se muestra una vista SVL_DATASHARE_USAGE_PRODUCER.

```
SELECT DISTINCT request_type
FROM svl_datashare_usage_producer
WHERE object_name LIKE 'tickit%';

   request_type
 ------------------
   "GET RELATION"
```

# SVL_FEDERATED_QUERY

Utilice la vista SVL_FEDERATED_QUERY para ver información sobre una llamada a consulta federada.

SVL_FEDERATED_QUERY es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_EXTERNAL_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | El ID del usuario que ejecuta la consulta. |
| xid | bigint | El ID de la transacción. |
| pid | entero | El ID del proceso principal que ejecuta la consulta. |
| consulta | entero | El ID de consulta de una llamada federada. |
| sourcetype | carácter (32) | El tipo de origen de llamada federada; por ejemplo “PG”. |
| recordtime | marca de tiempo | La hora a la que se envía una consulta para la federación. Se utiliza UTC. |
| querytext | carácter (4000) | La cadena de consulta enviada al motor PostgreSQL remoto para su ejecución. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| num_rows | bigint | El número de filas devueltas por la consulta federada. |
| num_bytes | bigint | El número de bytes devueltos por la consulta federada. |
| duración | bigint | El tiempo (microsegundos) empleado en recuperar filas de llamadas de cursor. Es el tiempo dedicado a ejecutar la consulta federada, así como a recuperar los resultados. |

Consultas de ejemplo

Para mostrar información acerca de las llamadas a consultas federadas, ejecute la siguiente consulta.

```
select query, trim(sourcetype) as type, recordtime, trim(querytext) as "PG Subquery"
 from svl_federated_query where query = 4292;

 query | type |         recordtime         |                         pg subquery
-------+------+----------------------------
+-------------------------------------------------------------
  4292 | PG   | 2020-03-27 04:29:58.485126 | SELECT "level" FROM functional.employees
 WHERE ("level" >= 6)
(1 row)
```

## SVL_MULTI_STATEMENT_VIOLATIONS

Utilice la vista SVL_MULTI_STATEMENT_VIOLATIONS para obtener un registro completo de todos los comandos SQL que se ejecutan en el sistema e infringen las restricciones de bloque de transacciones.

Las infracciones se producen cuando ejecuta cualquiera de los siguientes comandos SQL que Amazon Redshift restringe dentro de un bloque de transacción o solicitudes de varias instrucciones:

- CREATE DATABASE

- DROP DATABASE
- ALTER TABLE APPEND
- CREATE EXTERNAL TABLE
- DROP EXTERNAL TABLE
- RENAME EXTERNAL TABLE
- ALTER EXTERNAL TABLE
- CREATE TABLESPACE
- DROP TABLESPACE
- CREATE LIBRARY
- DROP LIBRARY
- REBUILDCAT
- INDEXCAT
- REINDEX DATABASE
- VACUUM
- GRANT
- COPY

> **Note**
>
> Si hay entradas en esta vista, cambie las aplicaciones y los scripts SQL correspondientes. Se recomienda cambiar el código de la aplicación para mover el uso de estos comandos SQL restringidos fuera del bloque de transacciones. Si necesita asistencia adicional, contacte con AWS Support.

SVL_MULTI_STATEMENT_VIOLATIONS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_QUERY_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | El ID del usuario que causó la infracción. |
| base de datos | character (32) | El nombre de la base de datos a la que estaba conectado el usuario. |
| cmdname | character(20) | El nombre del comando que no se puede ejecutar dentro de una solicitud de bloque de transacción o de varias instrucciones. Por ejemplo, CREATE DATABASE, DROP DATABASE, ALTER TABLE APPEND, CREATE EXTERNAL TABLE, DROP EXTERNAL TABLE, RENAME EXTERNAL TABLE, ALTER EXTERNAL TABLE, CREATE LIBRARY, DROP LIBRARY, REBUILDCAT, INDEXCAT, REINDEX DATABASE, VACUUM, GRANT en recursos externos, CLUSTER, COPY, CREATE TABLESPACE y DROP TABLESPACE. |
| xid | bigint | El ID de la transacción asociado a la instrucción. |
| pid | entero | El ID del proceso para la instrucción. |
| etiqueta | character(320) | Ya sea el nombre del archivo utilizado para ejecutar la consulta o una etiqueta definida con un comando SET QUERY_GROUP. Si la consulta no se basa en archivos o si no se establece el parámetro QUERY_GROUP, el valor del campo queda en blanco. |
| starttime | marca de tiempo | La hora exacta en que la instrucción comenzó a ejecutarse, con 6 dígitos de precisión para las fracciones de segundos; por ejemplo: **`2009-06-12 11:29:19.131358`** |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| endtime | marca de tiempo | La hora exacta en que la instrucción terminó de ejecutarse, con 6 dígitos de precisión para las fracciones de segundos; por ejemplo: **2009-06-12 11:29:19.193640** |
| sequence | entero | Cuando una única instrucción tiene más de 200 caractere s, se registran filas adicionales para esa instrucción. Secuencia 0 es la primera fila, 1 es la segunda y, así, sucesivamente. |
| type | varchar (10) | El tipo de instrucción SQL: **QUERY**, **DDL** o **UTILITY**. |
| texto | character(200) | El texto SQL, en incrementos de 200 caracteres. El campo puede contener caracteres especiales como una barra inversa(\\) y una nueva línea (\n). |

Consulta de ejemplo

En la siguiente consulta se devuelven varias instrucciones que tienen infracciones.

```
select * from svl_multi_statement_violations order by starttime asc;

userid | database | cmdname | xid | pid | label | starttime | endtime | sequence | type
 | text
================================================================================================
1 | dev | CREATE DATABASE | 1034 | 5729 |label1 | ******** | ******* | 0 | DDL |
 create table c(b int);
1 | dev | CREATE DATABASE | 1034 | 5729 |label1 | ******** | ******* | 0 | UTILITY |
 create database b;
1 | dev | CREATE DATABASE | 1034 | 5729 |label1 | ******** | ******* | 0 | UTILITY |
 COMMIT
...
```

## SVL_MV_REFRESH_STATUS

La vista SVL_MV_REFRESH_STATUS contiene una fila para la actividad de actualización de las vistas materializadas.

Para obtener más información acerca de las vistas materializadas, consulte Vistas materializadas en Amazon Redshift.

SVL_MV_REFRESH_STATUS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_MV_REFRESH_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| db_name | char(128) | Base de datos que contiene la vista materializada. |
| userid | bigint | El ID del usuario que realizó la actualización. |
| schema_name | char(128) | Esquema de la vista materializada. |
| mv_name | char(128) | El nombre de la vista materializada. |
| xid | bigint | El ID de transacción de la actualización. |
| starttime | marca de tiempo | La hora de inicio de la actualización. |
| endtime | marca de tiempo | La hora de finalización de la actualización. |
| status | texto | El estado de la actualización. Entre los valores de ejemplo se incluyen:<br>• La actualización ha actualizado gradualmente la máquina virtual con éxito<br>Si se trata de una vista materializada para streaming, el mensaje podría tener calificadores adicionales relativos al número de registros. Estos incluyen los siguientes: |

| Nombre de la columna | Tipo de datos: | Descripción |
| --- | --- | --- |
| | | • La transmisión no devolvió datos nuevos: no se recuperaron registros.<br>• Se omitieron todos los registros recibidos de la transmisión: se recuperaron los registros, pero, debido a errores, se omitieron todos.<br>• Se omitieron algunos registros de la transmisión: se recuperaron los registros, pero, debido a errores, se omitieron algunos.<br><br>Si no hay calificadores, se recuperó al menos un registro y todos los registros están disponibles en la vista materializada. Queda un posible calificador:<br><br>• La transmisión puede contener más datos: la actualización finalizó antes de que Amazon Redshift determinara que no había más registros que consumir. La transmisión puede estar actualizada, pero Amazon Redshift no la ha confirmado.<br>• La actualización ha vuelto a calcular la máquina virtual desde cero<br>• La actualización ha actualizado gradualmente la máquina virtual de forma parcial hacia una transacción activa<br>• La máquina virtual ya se ha actualizado<br>• Error al actualizar. Se ha cambiado el nombre de una columna de la tabla base<br>• Error al actualizar. Se ha cambiado un tipo de columna de la tabla base<br>• Error al actualizar. Se cambió el nombre de una tabla base |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | • La actualización produjo un error debido a un error interno<br>• Error al actualizar. Se ha eliminado una columna de la tabla base<br>• Error al actualizar. Se ha cambiado el nombre del esquema de la máquina virtual<br>• Error al actualizar. No se ha encontrado la máquina virtual<br>• Actualización automática anulada debido a una carga de trabajo excesiva del usuario<br>• Error al actualizar. Infracción de aislamiento serializable |
| refresh_type | char(32) | La definición del tipo de actualización. Entre los valores de ejemplo, se incluyen Manual y Auto. |

Consulta de ejemplo

Para ver el estado de actualización de las vistas materializadas, ejecute la siguiente consulta.

```
select * from svl_mv_refresh_status;
```

Esta consulta devuelve el siguiente ejemplo de salida:

```
 db_name | userid |  schema   |  name   |  xid  |         starttime          |
       endtime           |                        status                        |
 refresh_type
---------+--------+-----------+---------+-------+----------------------------
+----------------------------+-------------------------------------------------------
+-----------------
 dev     |    169 | mv_schema | mv_test |  6640 | 2020-02-14 02:26:53.497935 |
 2020-02-14 02:26:53.556156 | Refresh successfully recomputed MV from scratch     |
 Manual
```

```
dev     |    166 | mv_schema | mv_test |  6517 | 2020-02-14 02:26:39.287438 |
2020-02-14 02:26:39.349539 | Refresh successfully updated MV incrementally       |
Auto
dev     |    162 | mv_schema | mv_test |  6388 | 2020-02-14 02:26:27.863426 |
2020-02-14 02:26:27.918307 | Refresh successfully recomputed MV from scratch     |
Manual
dev     |    161 | mv_schema | mv_test |  6323 | 2020-02-14 02:26:20.020717 |
2020-02-14 02:26:20.080002 | Refresh successfully updated MV incrementally       |
Auto
dev     |    161 | mv_schema | mv_test |  6301 | 2020-02-14 02:26:05.796146 |
2020-02-14 02:26:07.853986 | Refresh successfully recomputed MV from scratch     |
Manual
dev     |    153 | mv_schema | mv_test |  6024 | 2020-02-14 02:25:18.762335 |
2020-02-14 02:25:20.043462 | MV was already updated                              |
Manual
dev     |    143 | mv_schema | mv_test |  5557 | 2020-02-14 02:24:23.100601 |
2020-02-14 02:24:23.100633 | MV was already updated                              |
Manual
dev     |    141 | mv_schema | mv_test |  5447 | 2020-02-14 02:23:54.102837 |
2020-02-14 02:24:00.310166 | Refresh successfully updated MV incrementally       |
Auto
dev     |      1 | mv_schema | mv_test |  5329 | 2020-02-14 02:22:26.328481 |
2020-02-14 02:22:28.369217 | Refresh successfully recomputed MV from scratch     |
Auto
dev     |    138 | mv_schema | mv_test |  5290 | 2020-02-14 02:21:56.885093 |
2020-02-14 02:21:56.885098 | Refresh failed. MV was not found                    |
Manual
```

## SVL_QERROR

La vista SVL_QERROR está en desuso.

## SVL_QLOG

La vista SVL_QLOG tiene un registro de todas las consultas que se ejecutan para la base de datos.

Amazon Redshift crea la vista SVL_QLOG como un subconjunto legible de información de la tabla STL_QUERY. Utilice esta tabla para encontrar el ID de consulta correspondiente a una consulta recientemente ejecutada o para ver cuánto demora en completarse una consulta.

SVL_QLOG es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_QUERY_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. Puede usar este ID para combinar distintas vistas y tablas del sistema. |
| xid | bigint | ID de transacción. |
| pid | entero | ID del proceso asociado a la consulta. |
| starttime | marca de tiempo | Hora exacta en que la instrucción comenzó a ejecutarse, con seis dígitos de precisión para las fracciones de segundos; por ejemplo: **2009-06-12 11:29:19.131358** |
| endtime | marca de tiempo | Hora exacta en que la instrucción terminó de ejecutarse, con seis dígitos de precisión para las fracciones de segundos; por ejemplo: **2009-06-12 11:29:19.193640** |
| elapsed | bigint | Tiempo que demoró la consulta en ejecutarse (en microsegundos). |
| aborted | entero | Si el sistema detuvo la consulta o el usuario la canceló, esta columna tendrá el valor **1**. Si la consulta se ejecutó hasta su finalización, esta columna tendrá el valor **0**. Las consultas que se cancelaron por motivos de administración |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | de cargas de trabajo y que se reiniciaron después también tienen un valor **1** en esta columna. |
| etiqueta | character(320) | Ya sea el nombre del archivo utilizado para ejecutar la consulta o una etiqueta definida con un comando SET QUERY_GROUP. Si la consulta no se basa en archivos o si no se establece el parámetro QUERY_GROUP, el valor del campo es `default`. |
| subcadena | character (60) | Texto de consulta truncado. |
| source_query | entero | Si la consulta utilizó el almacenamiento en caché de los resultados, es el ID de la consulta de origen de los resultados almacenados en caché. Si no se utilizó el almacenamiento en caché, el valor de este campo es `NULL`. |
| concurrency_scaling_status_txt | texto | Una descripción que indica si la consulta se ejecutó en el clúster principal o en un clúster de escalado de simultaneidad. |
| from_sp_call | entero | Si la consulta se llamó desde un procedimiento almacenado, el ID de la consulta de la llamada del procedimiento. Si la consulta no se ejecutó como parte de un procedimiento almacenado, este campo es `NULL`. |

Consultas de ejemplo

El siguiente ejemplo devuelve el ID de la consulta, el tiempo de ejecución y el texto truncado de la consulta para las cinco consultas más recientes de base de datos ejecutadas por el usuario con `userid = 100`.

```
select query, pid, elapsed, substring from svl_qlog
where userid = 100
order by starttime desc
```

```
limit 5;

 query  |  pid  | elapsed  |           substring
--------+-------+----------+-----------------------------------------------
 187752 | 18921 | 18465685 | select query, elapsed, substring from svl_...
 204168 |  5117 |    59603 | insert into testtable values (100);
 187561 | 17046 |  1003052 | select * from pg_table_def where tablename...
 187549 | 17046 |  1108584 | select * from STV_WLM_SERVICE_CLASS_CONFIG
 187468 | 17046 |  5670661 | select * from pg_table_def where schemaname...
(5 rows)
```

El siguiente ejemplo devuelve el nombre del script SQL (columna LABEL) y el tiempo transcurrido en una consulta que se canceló (**`aborted=1`**):

```
select query, elapsed, trim(label) querylabel
from svl_qlog where aborted=1;

 query | elapsed  |       querylabel
-------+----------+-------------------------
    16 |  6935292 | alltickittablesjoin.sql
(1 row)
```

## SVL_QUERY_METRICS

La vista SVL_QUERY_METRICS muestra las métricas de consultas completadas. Esta vista se deriva de la tabla de sistema STL_QUERY_METRICS. Utilice los valores de esta vista como ayuda para determinar los valores de umbral destinados a definir las reglas de supervisión de consultas. Para obtener más información, consulte Reglas de monitoreo de consultas de WLM.

SVL_QUERY_METRICS es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que ejecutó la consulta que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| service_class | entero | ID para la cola de consultas de WLM (clase de servicio). Las colas de consultas se definen en la configuración de WLM. Las métricas solo se informan para las colas definidas por el usuario. Para obtener una lista de ID de clase de servicio, consulte ID de la clase de servicio WLM. |
| dimension | varchar (24) | Dimensión en que se informa la métrica. Los valores posibles son consulta, segmento y paso. |
| segment | entero | Número del segmento. Una consulta consta de distintos segmentos y cada segmento consta de uno o más pasos. Los segmentos de una consulta se ejecutan en paralelo. Cada segmento se ejecuta en un solo proceso. Si el valor del segmento es 0, los valores métricos del segmento se acumulan hasta el nivel consulta. |
| paso | entero | ID para el tipo de paso que se realizó. La descripción para el tipo de paso se muestra en la columna `step_label` . |
| step_label | varchar (30) | Tipo de paso realizado. |
| query_cpu_time | bigint | Tiempo de CPU que utilizó la consulta, en segundos. El tiempo de CPU difiere del tiempo de ejecución de la consulta. |
| query_blocks_read | bigint | Número de bloques de 1 MB leídos por la consulta. |
| query_execution_time | bigint | Tiempo de ejecución transcurrido para una consulta, en segundos. El tiempo de ejecución no incluye el tiempo de |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | espera empleado en una cola. Ver `query_queue_time` para el tiempo en cola. |
| query_cpu_usage_percent | bigint | Porcentaje de capacidad de la CPU que utilizó la consulta. |
| query_temp_blocks_to_disk | bigint | La cantidad de espacio en el disco utilizada por una consulta para grabar los resultados intermedios, en MB. |
| segment_execution_time | bigint | Tiempo de ejecución transcurrido para un segmento único, en segundos. |
| cpu_skew | numeric (38,2) | La proporción entre el uso máximo de la CPU para cualquier sector y el uso promedio de la CPU para todos los sectores. La métrica se define a nivel segmento. |
| io_skew | numeric (38,2) | La proporción entre la lectura máxima de bloques (E/S) para cualquier sector y la lectura promedio de bloques para todos los sectores. |
| scan_row_count | bigint | La cantidad de filas en un paso del examen. La cantidad de filas es el número total de filas emitidas antes de filtrar las filas marcadas para eliminación (filas fantasma) y antes de aplicar los filtros de consulta definidos por el usuario. |
| join_row_count | bigint | La cantidad de filas procesadas en un paso de combinación. |
| nested_loop_join_row_count | bigint | El número de filas en una combinación de bucle anidado. |
| return_row_count | bigint | La cantidad de filas devueltas por la consulta. |
| spectrum_scan_row_count | bigint | El número de filas examinadas por Amazon Redshift Spectrum en Amazon S3. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| spectrum_ scan_size_mb | bigint | La cantidad de datos, en MB, examinados por Amazon Redshift Spectrum en Amazon S3. |
| query_que ue_time | bigint | Cantidad de tiempo en segundos que la consulta estuvo en cola. |

## SVL_QUERY_METRICS_SUMMARY

La vista SVL_QUERY_METRICS_SUMMARY muestra los valores máximos de métricas de consultas completadas. Esta vista se deriva de la tabla de sistema STL_QUERY_METRICS. Utilice los valores de esta vista como ayuda para determinar los valores de umbral destinados a definir las reglas de supervisión de consultas. Para obtener más información sobre las reglas y métricas para monitorear consultas para Amazon Redshift, consulte Reglas de monitoreo de consultas de WLM.

SVL_QUERY_METRICS_SUMMARY es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que ejecutó la consulta que generó la entrada. |
| consulta | entero | ID de la consulta. La columna de consulta puede usarse para combinar otras vistas y tablas del sistema. |
| service_class | entero | ID para la cola de consultas de WLM (clase de servicio). Las colas de consultas se definen en la configuración de WLM. Las métricas solo se informan para las colas definidas por |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | el usuario. Para obtener una lista de ID de clase de servicio, consulte ID de la clase de servicio WLM. |
| query_cpu_time | bigint | Tiempo de CPU que utilizó la consulta, en segundos. El tiempo de CPU difiere del tiempo de ejecución de la consulta. |
| query_blocks_read | bigint | Número de bloques de 1 MB leídos por la consulta. |
| query_execution_time | bigint | Tiempo de ejecución transcurrido para una consulta, en segundos. El tiempo de ejecución no incluye el tiempo de espera empleado en una cola. |
| query_cpu_usage_percent | numeric (38,2) | Porcentaje de capacidad de la CPU que utilizó la consulta. |
| query_temp_blocks_to_disk | bigint | La cantidad de espacio en el disco utilizada por una consulta para grabar los resultados intermedios, en MB. |
| segment_execution_time | bigint | Tiempo de ejecución transcurrido para un segmento único, en segundos. |
| cpu_skew | numeric (38,2) | La proporción entre el uso máximo de la CPU para cualquier sector y el uso promedio de la CPU para todos los sectores. La métrica se define a nivel segmento. |
| io_skew | numeric (38,2) | La proporción entre la lectura máxima de bloques (E/S) para cualquier sector y la lectura promedio de bloques para todos los sectores. |
| scan_row_count | bigint | La cantidad de filas en un paso del examen. La cantidad de filas es el número total de filas emitidas antes de filtrar las filas marcadas para eliminación (filas fantasma) y antes de aplicar los filtros de consulta definidos por el usuario. |
| join_row_count | bigint | La cantidad de filas procesadas en un paso de combinación. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| nested_loop_join_row_count | bigint | El número de filas en una combinación de bucle anidado. |
| return_row_count | bigint | La cantidad de filas devueltas por la consulta. |
| spectrum_scan_row_count | bigint | El número de filas examinadas por Amazon Redshift Spectrum en Amazon S3. |
| spectrum_scan_size_mb | bigint | La cantidad de datos, en MB, examinados por Amazon Redshift Spectrum en Amazon S3. |
| query_queue_time | bigint | Cantidad de tiempo en segundos que la consulta estuvo en cola. |

## SVL_QUERY_QUEUE_INFO

Resume los detalles de las consultas que pasan tiempo en una cola de consultas o en una cola de confirmación de Workload Management (WLM, Administración de cargas de trabajo).

La vista SVL_QUERY_QUEUE_INFO filtra las consultas realizadas por el sistema y muestra solo las realizadas por un usuario.

La vista SVL_QUERY_QUEUE_INFO resume información acerca de las tablas de sistema STL_QUERY, STL_WLM_QUERY y STL_COMMIT_STATS.

SVL_QUERY_QUEUE_INFO solo es visible para los superusuarios. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| base de datos | texto | El nombre de la base de datos al que estaba conectado el usuario cuando se emitió la consulta. |
| consulta | entero | ID de la consulta. |
| xid | bigint | ID de transacción. |
| userid | entero | ID del usuario que generó la consulta. |
| querytxt | texto | Primeros 100 caracteres del texto de la consulta. |
| queue_start_time | marca de tiempo | Hora en UTC en la que ingresó la consulta a la cola de WLM. |
| exec_start_time | marca de tiempo | Tiempo en UTC en que comenzó la ejecución de la consulta. |
| service_class | entero | ID para la clase de servicio. Las clases de servicios están definidas en el archivo de configuración de WLM. |
| slots | entero | Cantidad de slots de consulta de WLM. |
| queue_elapsed | bigint | Tiempo que estuvo la consulta en la cola de WLM (en segundos). |
| exec_elapsed | bigint | Tiempo transcurrido de la consulta en ejecución (en segundos). |
| wlm_total_elapsed | bigint | Tiempo en que la consulta estuvo en cola de WLM (queue_elapsed) más el tiempo en que la consulta estuvo en ejecución (exec_elapsed). |
| commit_queue_elapsed | bigint | Tiempo que estuvo la consulta en la cola de confirmaciones (en segundos). |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| commit_exec_time | bigint | Tiempo que estuvo la consulta en la operación de confirmación (en segundos). |
| service_class_name | character(64) | El nombre de la clase de servicio. |

Consultas de ejemplo

En el siguiente ejemplo, se muestra el tiempo en que las consultas estuvieron en colas de WLM.

```
select query, service_class, queue_elapsed, exec_elapsed, wlm_total_elapsed
from svl_query_queue_info
where wlm_total_elapsed > 0;

  query  | service_class | queue_elapsed | exec_elapsed | wlm_total_elapsed
---------+---------------+---------------+--------------+------------------
 2742669 |             6 |             2 |          916 |               918
 2742668 |             6 |             4 |          197 |               201
(2 rows)
```

## SVL_QUERY_REPORT

Amazon Redshift crea la vista SVL_QUERY_REPORT de una función UNION de una cantidad de tablas de sistema STL de Amazon Redshift para proporcionar información acerca de los pasos completados de la consulta.

Esta vista desglosa la información relacionada con las consultas completadas por sector y por paso, lo que puede ayudar con el nodo de solución de problemas y con los problemas del sector en el clúster de Amazon Redshift.

SVL_QUERY_REPORT es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. Se puede usar para combinar distintas vistas y tablas del sistema. |
| slice | entero | Sector de datos donde se ejecutó el paso. |
| segment | entero | Número del segmento.<br><br>Una consulta consta de distintos segmentos y cada segmento consta de uno o más pasos. Los segmentos de una consulta se ejecutan en paralelo. Cada segmento se ejecuta en un solo proceso. |
| paso | entero | Paso de la consulta que se completó. |
| start_time | marca de tiempo | Hora exacta en UTC en que un segmento comenzó a ejecutarse, con 6 dígitos de precisión para las fracciones de segundos. Por ejemplo: **`2012-12-12 11:29:19.131358`** . |
| end_time | marca de tiempo | Hora exacta en UTC en que un segmento finalizó la ejecución, con 6 dígitos de precisión para las fracciones de segundos. Por ejemplo: **`2012-12-12 11:29:19.131467`** . |
| elapsed_time | bigint | Tiempo (en microsegundos) que demoró el segmento en ejecutarse. |
| rows | bigint | Cantidad de filas producidas por el paso (por sector). Este número representa la cantidad de filas para el sector que son el resultado de la ejecución del paso, no la cantidad de filas recibidas ni procesadas por el paso. En otras palabras, es la cantidad de filas que sobreviven al paso y que pasan al siguiente. |
| bytes | bigint | Cantidad de bytes producidos por el paso (por sector). |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| etiqueta | char (256) | Etiqueta del paso, que se compone del nombre de un paso de la consulta y, si corresponde, el ID y el nombre de la tabla (por ejemplo, `scan tbl=100448 name =user` ). Los ID de tabla de tres dígitos suelen referirse a exámenes de tablas transitorias. Cuando vea `tbl=0`, por lo general se refiere a un examen de un valor constante. |
| is_diskbased | character (1) | Indica si este paso de la consulta se ejecutó como una operación basada en el disco: true (**t**) (verdadero) o false (**f**) (falso). Solo determinados pasos, como hash, ordenar y agregar pasos, pueden ir al disco. Muchos tipos de pasos se realizan siempre en la memoria. |
| workmem | bigint | Cantidad de memoria funcional (en bytes) asignada a este paso de la consulta. Este valor es el umbral query_working_mem asignado para el uso durante la ejecución, no la cantidad de memoria que realmente se usó. |
| is_rrscan | character (1) | Si es true (**t**), indica que se utilizó un examen de rango restringido en el paso. |
| is_delayed_scan | character (1) | Si es true (**t**), indica que en el paso se utilizó un examen diferido. |
| rows_pre_filter | bigint | Para exámenes de tablas permanentes, la cantidad total de filas emitidas antes de filtrar las filas marcadas para eliminación (filas fantasma) y antes de aplicar los filtros de consulta definidos por el usuario. |

Consultas de ejemplo

En la siguiente consulta, se demuestra el sesgo de datos de las filas devueltas para la consulta con el ID de consulta 279. Utilice esta consulta para determinar si los datos de la base de datos están distribuidos uniformemente en los sectores en el clúster del data warehouse:

```
select query, segment, step, max(rows), min(rows),
```

```
case when sum(rows) > 0
then ((cast(max(rows) -min(rows) as float)*count(rows))/sum(rows))
else 0 end
from svl_query_report
where query = 279
group by query, segment, step
order by segment, step;
```

Esta consulta debe devolver datos similares al siguiente ejemplo de salida:

```
query | segment | step |   max    |   min    |         case
------+---------+------+----------+----------+----------------------
279 |       0 |    0 | 19721687 | 19721687 |                    0
279 |       0 |    1 | 19721687 | 19721687 |                    0
279 |       1 |    0 |   986085 |   986084 | 1.01411202804304e-06
279 |       1 |    1 |   986085 |   986084 | 1.01411202804304e-06
279 |       1 |    4 |   986085 |   986084 | 1.01411202804304e-06
279 |       2 |    0 |  1775517 |   788460 |     1.00098637606408
279 |       2 |    2 |  1775517 |   788460 |     1.00098637606408
279 |       3 |    0 |  1775517 |   788460 |     1.00098637606408
279 |       3 |    2 |  1775517 |   788460 |     1.00098637606408
279 |       3 |    3 |  1775517 |   788460 |     1.00098637606408
279 |       4 |    0 |  1775517 |   788460 |     1.00098637606408
279 |       4 |    1 |  1775517 |   788460 |     1.00098637606408
279 |       4 |    2 |        1 |        1 |                    0
279 |       5 |    0 |        1 |        1 |                    0
279 |       5 |    1 |        1 |        1 |                    0
279 |       6 |    0 |       20 |       20 |                    0
279 |       6 |    1 |        1 |        1 |                    0
279 |       7 |    0 |        1 |        1 |                    0
279 |       7 |    1 |        0 |        0 |                    0
(19 rows)
```

## SVL_QUERY_SUMMARY

Utilice la vista SVL_QUERY_SUMMARY para encontrar información general acerca de la ejecución de una consulta.

La vista SVL_QUERY_SUMMARY tiene un subconjunto de datos de la vista SVL_QUERY_REPORT. Tenga en cuenta que la información en SVL_QUERY_SUMMARY proviene de todos los nodos.

> **Note**
>
> La vista SVL_QUERY_SUMMARY solo contiene información relacionada con las consultas realizadas por Amazon Redshift, no otros comandos de utilidades ni de DDL. Para obtener una lista completa e información acerca de todas las instrucciones realizadas por Amazon Redshift, incluidos los comandos de DDL y de utilidades, puede consultar la vista SVL_STATEMENTTEXT.

SVL_QUERY_SUMMARY es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Para obtener más información sobre SVCS_QUERY_SUMMARY, consulte SVCS_QUERY_SUMMARY.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| consulta | entero | ID de la consulta. Se puede usar para combinar distintas vistas y tablas del sistema. |
| stm | entero | Secuencia: un conjunto de segmentos simultáneos en una consulta. Una consulta tiene una o más secuencias. |
| seg | entero | Número del segmento. Una consulta consta de distintos segmentos y cada segmento consta de uno o más pasos. Los segmentos de una consulta se ejecutan en paralelo. Cada segmento se ejecuta en un solo proceso. |
| paso | entero | Paso de la consulta que se ejecutó. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| maxtime | bigint | Cantidad máxima de tiempo en que se ejecuta el paso (en microsegundos). |
| avgtime | bigint | Tiempo promedio en que se ejecuta el paso (en microsegundos). |
| rows | bigint | Cantidad de datos involucrados en el paso de la consulta. |
| bytes | bigint | Cantidad de bytes de datos involucrados en el paso de la consulta. |
| rate_row | double precision | Proporción de ejecución de la consulta por fila. |
| rate_byte | double precision | Proporción de ejecución de la consulta por byte. |
| etiqueta | texto | Etiqueta del paso, que se compone del nombre de un paso de la consulta y, si corresponde, el ID y el nombre de la tabla (por ejemplo, scan tbl=100448 name =user). Los ID de tabla de tres dígitos suelen referirse a exámenes de tablas transitorias. Cuando vea `tbl=0`, por lo general se refiere a un examen de un valor constante. |
| is_diskbased | character (1) | Indica si este paso de la consulta se realizó como una operación basada en el disco en algún nodo del clúster: true (**t**) (verdadero) o false (**f**) (falso). Solo determinados pasos, como hash, ordenar y agregar pasos, pueden ir al disco. Muchos tipos de pasos se realizan siempre en la memoria. |
| workmem | bigint | Cantidad de memoria funcional (en bytes) asignada a este paso de la consulta. |
| is_rrscan | character (1) | Si es true (**t**), indica que se utilizó un examen de rango restringido en el paso. El valor predeterminado es false (**f**). |
| is_delayed_scan | character (1) | Si es true (**t**), indica que en el paso se utilizó un examen diferido. El valor predeterminado es false (**f**). |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| rows_pre_filter | bigint | Para exámenes de tablas permanentes, la cantidad total de filas emitidas antes de filtrar las filas marcadas para eliminación (filas fantasma). |

Consultas de ejemplo

Vista de la información de procesamiento del paso de una consulta

En la siguiente consulta, se muestra la información básica de procesamiento para cada paso de la consulta 87:

```
select query, stm, seg, step, rows, bytes
from svl_query_summary
where query = 87
order by query, seg, step;
```

Esta consulta recupera la información de procesamiento acerca de la consulta 87, tal como se muestra en el siguiente ejemplo de salida:

```
 query | stm | seg | step |  rows  |  bytes
-------+-----+-----+------+--------+---------
87     |   0 |   0 |    0 |     90 |    1890
87     |   0 |   0 |    2 |     90 |     360
87     |   0 |   1 |    0 |     90 |     360
87     |   0 |   1 |    2 |     90 |    1440
87     |   1 |   2 |    0 | 210494 | 4209880
87     |   1 |   2 |    3 |  89500 |       0
87     |   1 |   2 |    6 |      4 |      96
87     |   2 |   3 |    0 |      4 |      96
87     |   2 |   3 |    1 |      4 |      96
87     |   2 |   4 |    0 |      4 |      96
87     |   2 |   4 |    1 |      1 |      24
87     |   3 |   5 |    0 |      1 |      24
87     |   3 |   5 |    4 |      0 |       0
(13 rows)
```

Determinar qué pasos de la consulta se vertieron en el disco

En la siguiente consulta, se muestra si alguno de los pasos de la consulta con el ID de consulta 1025 (ver la vista SVL_QLOG para saber cómo obtener el ID de una consulta) se vertieron al disco o si la consulta se ejecutó en su totalidad en la memoria:

```
select query, step, rows, workmem, label, is_diskbased
from svl_query_summary
where query = 1025
order by workmem desc;
```

Esta consulta devuelve el siguiente ejemplo de salida:

```
query| step|  rows  |  workmem   |  label        | is_diskbased
-----+-----+--------+-----------+---------------+--------------
1025 |  0  |16000000|  141557760 |scan tbl=9     | f
1025 |  2  |16000000|  135266304 |hash tbl=142   | t
1025 |  0  |16000000|  128974848 |scan tbl=116536| f
1025 |  2  |16000000|  122683392 |dist           | f
(4 rows)
```

Al examinar los valores para IS_DISKBASED, puede ver qué pasos de la consulta pasaron al disco. Para la consulta 1025, el paso hash se ejecutó en el disco. Los pasos que se pueden ejecutar en el disco incluyen los pasos hash, aggr y sort. Para ver únicamente los pasos de la consulta basados en el disco, agregue la cláusula **`and is_diskbased = 't'`** a la instrucción SQL del ejemplo anterior.

## SVL_RESTORE_ALTER_TABLE_PROGRESS

Utilice SVL_RESTORE_ALTER_TABLE_PROGRESS para monitorear el progreso de la migración de cada tabla del clúster durante un cambio de tamaño clásico en los nodos RA3. Captura el rendimiento histórico de la migración de datos durante la operación de cambio de tamaño. Para obtener más información sobre el cambio de tamaño clásico a los nodos RA3, consulte Cambio de tamaño clásico.

SVL_RESTORE_ALTER_TABLE_PROGRESS solo es visible para los superusuarios. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_RESTORE_LOG. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

> **Note**
>
> Las filas con un progreso de `100.00%` o `ABORTED` se eliminan después de 7 días. Las filas de las tablas eliminadas durante o después de un cambio de tamaño clásico pueden seguir apareciendo en SVL_RESTORE_ALTER_TABLE_PROGRESS.

### Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| tbl | entero | ID de la tabla. |
| avance | char(32) | Estado del progreso de redistribución de la tabla. Los valores posibles son porcentajes desde `0.00%` hasta `100.00%` y el mensaje `ABORTED`. `ABORTED` significa que la redistribución se detuvo sin terminar y el motivo se explica en la columna `message`. |
| mensaje | char (256) | El mensaje asociado con el progreso de la redistribución de la tabla. |

### Consulta de ejemplo

La siguiente consulta devuelve consultas en ejecución y en cola.

```
select * from svl_restore_alter_table_progress;

tbl     | progress |                          message
--------+----------+-------------------------------------------------------------
105614  | ABORTED  | Abort:Table no longer contains the prior dist key column.
105610  | ABORTED  | Abort:Table no longer contains the prior dist key column.
105594  | 0.00%    | Table waiting for alter diststyle conversion.
105602  | ABORTED  | Abort:Table no longer contains the prior dist key column.
105606  | ABORTED  | Abort:Table no longer contains the prior dist key column.
105598  | 100.00%  | Restored to distkey successfully.
```

# SVL_S3LIST

Utilice la vista SVL_S3LIST para obtener detalles acerca de las consultas de Amazon Redshift Spectrum en el nivel de segmento.

SVL_S3LIST es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> SVL_S3LIST solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_EXTERNAL_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| consulta | entero | El ID de la consulta. |
| segment | entero | El número del segmento. Una consulta consta de varios segmentos. |
| node | entero | El número de nodo. |
| slice | entero | El sector de datos contra el que se ejecutó un segmento en particular. |
| eventtime | marca de tiempo | La hora en UTC a la que se registra el evento. |
| bucket | texto | El nombre del bucket de Amazon S3. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| prefix | texto | El prefijo de la ubicación del bucket de Amazon S3. |
| recursive | char(1) | Si se realiza un análisis recursivo de las subcarpetas. |
| retrieved_files | entero | El número de archivos mostrados. |
| max_file_size | bigint | El tamaño de archivo máximo entre los archivos mostrados. |
| avg_file_size | double precision | El tamaño de archivo medio entre los archivos mostrados. |
| generated_splits | entero | El número de divisiones de archivo. |
| avg_split_length | double precision | La longitud media de las divisiones de archivos en bytes. |
| duración | bigint | La duración del listado de archivos, en microsegundos. |

Consulta de ejemplo

En el siguiente ejemplo, se consulta SVL_S3LIST para obtener la última consulta que se va a ejecutar.

```
select *
from svl_s3list
where query = pg_last_query_id()
order by query,segment;
```

## SVL_S3LOG

Utilice la vista SVL_S3LOG para obtener detalles acerca de las consultas de Amazon Redshift Spectrum en el nivel de segmento y de sector del nodo.

SVL_S3LOG es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> SVL_S3LOG solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_EXTERNAL_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| pid | entero | El ID del proceso. |
| consulta | entero | El ID de la consulta. |
| segment | entero | El número del segmento. Una consulta consta de distintos segmentos y cada segmento consta de uno o más pasos. |
| paso | entero | El paso de la consulta que se ejecutó. |
| node | entero | El número de nodo. |
| slice | entero | El sector de datos contra el que se ejecutó un segmento en particular. |
| eventtime | marca de tiempo | Hora en UTC en que comenzó a ejecutarse el paso. |
| mensaje | texto | Mensaje para la entrada del registro. |

Consulta de ejemplo

En el siguiente ejemplo, se consulta SVL_S3LOG para obtener la última consulta que se ejecutó.

```
select *
from svl_s3log
where query = pg_last_query_id()
order by query,segment,slice;
```

## SVL_S3PARTITION

Utilice la vista SVL_S3PARTITION para obtener detalles acerca de las particiones de Amazon Redshift Spectrum en el nivel de segmento y de sector del nodo.

SVL_S3PARTITION es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> SVL_S3PARTITION solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_EXTERNAL_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| consulta | entero | El ID de la consulta. |
| segment | entero | Un número de segmento. Una consulta consta de distintos segmentos y cada segmento consta de uno o más pasos. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| node | entero | El número de nodo. |
| slice | entero | El sector de datos contra el que se ejecutó un segmento en particular. |
| starttime | TIMESTAMP sin zona horaria | Hora en UTC en que comenzó a ejecutarse la operación de recorte de la partición. |
| endtime | TIMESTAMP sin zona horaria | Hora en UTC en que se completó la operación de recorte de la partición. |
| duración | bigint | Tiempo transcurrido (en microsegundos). |
| total_partitions | entero | Número total de particiones. |
| qualified_partitions | entero | Número de particiones calificadas. |
| assigned_partitions | entero | Número de particiones asignadas en el sector. |
| assignment | character | Tipo de asignación. |

Consulta de ejemplo

En el siguiente ejemplo, se obtienen los detalles de partición de la última consulta completada.

```
SELECT query, segment,
       MIN(starttime) AS starttime,
       MAX(endtime) AS endtime,
       datediff(ms,MIN(starttime),MAX(endtime)) AS dur_ms,
       MAX(total_partitions) AS total_partitions,
       MAX(qualified_partitions) AS qualified_partitions,
       MAX(assignment) as assignment_type
FROM svl_s3partition
```

```
WHERE query=pg_last_query_id()
GROUP BY query, segment
```

```
query | segment |           starttime           |           endtime           | dur_ms|
 total_partitions | qualified_partitions | assignment_type
------+---------+-------------------------------+-----------------------------+-------
+------------------+----------------------+---------------
99232 |       0 | 2018-04-17 22:43:50.201515    | 2018-04-17 22:43:54.674595  |  4473 |
       2526       |        334           | p
```

## SVL_S3PARTITION_SUMMARY

Utilice la vista SVL_S3PARTITION_SUMMARY para obtener un resumen de las particiones de consultas de Redshift Spectrum que se procesan en el nivel de segmento.

SVL_S3PARTITION_SUMMARY es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Para obtener más información sobre SVCS_S3PARTITION, consulte SVCS_S3PARTITION_SUMMARY.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| consulta | entero | El ID de la consulta. Puede usar este valor para combinar distintas vistas y tablas del sistema. |
| segment | entero | El número del segmento. Una consulta consta de varios segmentos. |
| assignment | char(1) | El tipo de asignación de partición entre nodos. |
| min_starttime | marca de tiempo | Hora en UTC a la que se inició el procesamiento de la partición. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| max_endtime | marca de tiempo | Hora en UTC a la que finalizó el procesamiento de la partición. |
| min_duration | bigint | Tiempo mínimo de procesamiento de la partición usado por un nodo para esta consulta (en microsegundos). |
| max_duration | bigint | Tiempo máximo de procesamiento de la partición usado por un nodo para esta consulta (en microsegundos). |
| avg_duration | bigint | Tiempo medio de procesamiento de la partición usado por un nodo para esta consulta (en microsegundos). |
| total_partitions | entero | Número total de particiones en una tabla externa. |
| qualified_partitions | entero | Número total de particiones aptas. |
| min_assigned_partitions | entero | Número mínimo de particiones asignadas en un nodo. |
| max_assigned_partitions | entero | Número máximo de particiones asignadas en un nodo. |
| avg_assigned_partitions | bigint | Número medio de particiones asignadas en un nodo. |

Consulta de ejemplo

En el siguiente ejemplo, se obtienen los detalles de análisis de partición para la última consulta completada.

```
select query, segment, assignment, min_starttime, max_endtime, min_duration,
 avg_duration
from svl_s3partition_summary
where query = pg_last_query_id()
order by query,segment;
```

## SVL_S3QUERY

Utilice la vista SVL_S3QUERY para obtener detalles acerca de las consultas de Amazon Redshift Spectrum en el nivel de segmento y de sector del nodo.

SVL_S3QUERY es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

> **Note**
>
> SVL_S3QUERY solo contiene consultas que se ejecutan en los clústeres aprovisionados principales. No contiene consultas que se ejecutan en clústeres de escalado simultáneo ni en espacios de nombres sin servidor. Para acceder a los planes de explicación de consultas ejecutadas tanto en los clústeres principales como en los clústeres de escalado simultáneo y en los espacios de nombres sin servidor, le recomendamos que utilice la vista de supervisión de SYS SYS_EXTERNAL_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | El ID del usuario que generó una entrada dada. |
| consulta | entero | El ID de la consulta. |
| segment | entero | Un número de segmento. Una consulta consta de distintos segmentos y cada segmento consta de uno o más pasos. |
| paso | entero | El paso de la consulta que se ejecutó. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| node | entero | El número de nodo. |
| slice | entero | El sector de datos contra el que se ejecutó un segmento en particular. |
| starttime | marca de tiempo | Hora en UTC en que comenzó a ejecutarse la consulta. |
| endtime | marca de tiempo | Hora en UTC en que se completó la ejecución de la consulta. |
| elapsed | entero | Tiempo transcurrido (en microsegundos). |
| external_table_name | char (136) | Formato interno del nombre de tabla externa para el paso de examen S3. |
| is_partitioned | char(1) | Si es true (`t`), el valor de esta columna indica que la tabla externa está particionada. |
| is_rrscan | char (1) | Si es true (`t`), el valor de esta columna indica que se aplicó un examen de rango restringido. |
| s3_scanned_rows | bigint | La cantidad de filas examinadas desde Amazon S3 y enviadas a la capa de Redshift Spectrum. |
| s3_scanned_bytes | bigint | La cantidad de bytes examinados desde Amazon S3 y enviados a la capa Redshift Spectrum. |
| s3query_returned_rows | bigint | La cantidad de filas devueltas desde la capa Redshift Spectrum hasta el clúster. |
| s3query_returned_bytes | bigint | La cantidad de bytes devueltos desde la capa Redshift Spectrum hasta el clúster. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| files | entero | La cantidad de filas que se procesaron para este paso de examen S3 en este sector. |
| splits | int | Número de divisiones procesadas en este sector. En el caso de los archivos de datos divisibles de gran tamaño (por ejemplo, archivos que superan los 512 MB), Redshift Spectrum intenta dividir los archivos en varias solicitudes de S3 para procesarlos en paralelo. |
| total_split_size | bigint | Tamaño total en bytes de las divisiones de este sector. |
| max_split_size | bigint | Tamaño máximo en bytes de las divisiones procesadas en este sector. |
| total_retries | entero | Cantidad total de reintentos para los archivos procesados. |
| max_retries | entero | Cantidad máxima de reintentos para un archivo individual procesado. |
| max_request_duration | entero | La duración máxima de una solicitud individual de Redshift Spectrum (en microsegundos). |
| avg_request_duration | double precision | La duración promedio de las solicitudes de Redshift Spectrum (en microsegundos). |
| max_request_parallelism | entero | La cantidad máxima de Redshift Spectrum pendiente en este sector para este paso de examen S3. |
| avg_request_parallelism | double precision | La cantidad promedio de solicitudes en paralelo de Redshift Spectrum en este sector para este paso de examen S3. |

Consulta de ejemplo

En el siguiente ejemplo, se obtienen los detalles del paso de análisis para la última consulta completada.

```
select query, segment, slice, elapsed, s3_scanned_rows, s3_scanned_bytes,
 s3query_returned_rows, s3query_returned_bytes, files
from svl_s3query
where query = pg_last_query_id()
order by query,segment,slice;
```

```
query | segment | slice | elapsed | s3_scanned_rows | s3_scanned_bytes |
 s3query_returned_rows | s3query_returned_bytes | files
------+---------+-------+---------+-----------------+------------------
+----------------------+-------------------------+------
 4587 |       2 |     0 |   67811 |               0 |                0 |
      0 |                      0 |     0
 4587 |       2 |     1 |  591568 |          172462 |         11260097 |
   8513 |                  170260 |     1
 4587 |       2 |     2 |  216849 |               0 |                0 |
      0 |                      0 |     0
 4587 |       2 |     3 |  216671 |               0 |                0 |
      0 |                      0 |     0
```

## SVL_S3QUERY_SUMMARY

Utilice la vista SVL_S3QUERY_SUMMARY para obtener un resumen de todas las consultas de Amazon Redshift Spectrum (consultas S3) que se ejecutaron en el sistema. SVL_S3QUERY_SUMMARY agrega detalles de la vista SVL_S3QUERY a nivel segmento.

SVL_S3QUERY_SUMMARY es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_EXTERNAL_QUERY_DETAIL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Para obtener más información sobre SVCS_S3QUERY_SUMMARY, consulte SVCS_S3QUERY_SUMMARY.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada dada. |
| consulta | entero | El ID de la consulta. Puede usar este valor para combinar distintas vistas y tablas del sistema. |
| xid | bigint | El ID de la transacción. |
| pid | entero | El ID del proceso. |
| segment | entero | El número del segmento. Una consulta consta de distintos segmentos y cada segmento consta de uno o más pasos. |
| paso | entero | El paso de la consulta que se ejecutó. |
| starttime | marca de tiempo | Hora en UTC en que comenzó a ejecutarse la consulta. |
| endtime | marca de tiempo | Hora en UTC en que se completó la consulta. |
| elapsed | entero | Tiempo que demoró la consulta en ejecutarse (en microsegundos). |
| aborted | entero | Si el sistema detuvo la consulta o el usuario la canceló, esta columna tendrá el valor **1**. Si la consulta se ejecutó hasta su finalización, esta columna tendrá el valor **0**. |
| external_table_name | char (136) | El formato interno del nombre del nombre externo de la tabla para el examen de la tabla externa. |
| file_format | character(16) | El formato de archivo de los datos de la tabla externa. |
| is_partitioned | char(1) | Si es true (**t**), el valor de esta columna indica que la tabla externa está particionada. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| is_rrscan | char (1) | Si es true (**t**), el valor de esta columna indica que se aplicó un examen de rango restringido. |
| is_nested | char(1) | Si es true (**t**), el valor de esta columna indica que se obtuvo acceso al tipo de datos de la columna anidada. |
| s3_scanned_rows | bigint | La cantidad de filas examinadas desde Amazon S3 y enviadas a la capa de Redshift Spectrum. |
| s3_scanned_bytes | bigint | La cantidad de bytes examinados desde Amazon S3 y enviados a la capa de Redshift Spectrum, según los datos comprimidos. |
| s3query_returned_rows | bigint | La cantidad de filas devueltas desde la capa Redshift Spectrum hasta el clúster. |
| s3query_returned_bytes | bigint | La cantidad de bytes devueltos desde la capa Redshift Spectrum hasta el clúster. Una gran cantidad de datos devueltos a Amazon Redshift podría afectar el rendimiento del sistema. |
| files | entero | La cantidad de filas que se procesaron para esta consulta de Redshift Spectrum. Una pequeña cantidad de archivos limita los beneficios del procesamiento en paralelo. |
| files_max | entero | Cantidad total de archivos procesados en un sector. |
| files_avg | entero | Cantidad media de archivos procesados en un sector. |
| splits | int | Número de divisiones procesadas para este segmento. Número de divisiones procesadas en este sector. En el caso de los archivos de datos divisibles de gran tamaño (por ejemplo, archivos que superan los 512 MB), Redshift Spectrum intenta dividir los archivos en varias solicitudes de S3 para procesarlos en paralelo. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| splits_max | int | Número máximo de divisiones procesadas en este sector. |
| splits_avg | int | Número medio de divisiones procesadas en este sector. |
| total_split_size | bigint | Tamaño total de todas las divisiones procesadas. |
| max_split_size | bigint | Tamaño máximo en bytes de las divisiones procesadas. |
| avg_split_size | bigint | Tamaño medio en bytes de las divisiones procesadas. |
| total_retries | entero | Cantidad total de reintentos para un archivo individual procesado. |
| max_retries | entero | Cantidad máxima de reintentos para cualquiera de los archivos procesados. |
| max_request_duration | entero | La duración máxima de una solicitud de un archivo individual (en microsegundos). Si las consultas tienen demasiado tiempo de ejecución, esto podría deberse a un cuello de botella. |
| avg_request_duration | double precision | La duración media de las solicitudes de archivo (en microsegundos). |
| max_request_parallelism | entero | La cantidad máxima de solicitudes en paralelo en un sector para esta consulta de Redshift Spectrum. |
| avg_request_parallelism | double precision | La cantidad promedio de solicitudes en paralelo en un sector para esta consulta de Redshift Spectrum. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| total_slowdown_count | bigint | La cantidad total de solicitudes de Amazon S3 con un error de procesamiento lento que se produjeron durante el análisis de la tabla externa. |
| max_slowdown_count | entero | La cantidad máxima de solicitudes de Amazon S3 con un error de procesamiento lento que se produjeron durante el análisis de la tabla externa en un sector. |

Consulta de ejemplo

En el siguiente ejemplo, se obtienen los detalles del paso de análisis para la última consulta completada.

```
select query, segment, elapsed, s3_scanned_rows, s3_scanned_bytes,
 s3query_returned_rows, s3query_returned_bytes, files
from svl_s3query_summary
where query = pg_last_query_id()
order by query,segment;
```

```
query | segment | elapsed | s3_scanned_rows | s3_scanned_bytes | s3query_returned_rows
 | s3query_returned_bytes | files
------+---------+---------+-----------------+------------------+----------------------
+-------------------------+------
 4587 |       2 |   67811 |               0 |                0 |                    0
 |                     0 |     0
 4587 |       2 |  591568 |          172462 |         11260097 |                 8513
 |                170260 |     1
 4587 |       2 |  216849 |               0 |                0 |                    0
 |                     0 |     0
 4587 |       2 |  216671 |               0 |                0 |                    0
 |                     0 |     0
```

## SVL_S3RETRIES

Utilice la vista SVL_S3RETRIES para obtener información acerca de por qué se produjeron errores en una consulta de Amazon Redshift Spectrum basada en Amazon S3.

SVL_S3RETRIES es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| consulta | entero | El ID de la consulta. |
| segment | entero | Número del segmento.<br><br>Una consulta consta de distintos segmentos y cada segmento consta de uno o más pasos. Los segmentos de una consulta se ejecutan en paralelo. Cada segmento se ejecuta en un solo proceso. |
| node | entero | El número de nodo. |
| slice | entero | El sector de datos contra el que se ejecutó un segmento en particular. |
| eventtime | TIMESTAMP sin zona horaria | Hora en UTC en que comenzó a ejecutarse el paso. |
| retries | entero | Número de reintentos de la consulta. |
| successful_fetches | entero | Número de veces que se devolvieron datos. |

| Nombre de la columna | Tipo de datos: | Descripción | | |
|---|---|---|---|---|
| file_size | bigint | Este tamaño del archivo en bytes. | | |
| ubicación | texto | La ubicación de la tabla. | | |
| mensaje | texto | Mensaje de error. | | |

Consulta de ejemplo

En el siguiente ejemplo, se recuperan datos sobre consultas de S3 con errores.

```
SELECT svl_s3retries.query, svl_s3retries.segment, svl_s3retries.node,
 svl_s3retries.slice, svl_s3retries.eventtime, svl_s3retries.retries,
svl_s3retries.successful_fetches, svl_s3retries.file_size,
 btrim((svl_s3retries."location")::text) AS "location",
 btrim((svl_s3retries.message)::text)
AS message FROM svl_s3retries;
```

## SVL_SPATIAL_SIMPLIFY

Puede consultar la vista del sistema SVL_SPATIAL_SIMPLIFY para obtener información acerca de los objetos de geometría espacial simplificada mediante el comando COPY. Cuando utiliza COPY en un shapefile, puede especificar las opciones de ingesta SIMPLIFY `tolerance`, SIMPLIFY AUTO y SIMPLIFY AUTO `max_tolerance`. El resultado de la simplificación se resume en la vista de sistema SVL_SPATIAL_SIMPLIFY.

Cuando se configura SIMPLIFY AUTO `max_tolerance`, esta vista contiene una fila para cada geometría que excedió el tamaño máximo. Cuando se configura SIMPLIFY `tolerance`, se almacena una fila para toda la operación COPY. Esta fila referencia el ID de consulta COPY y la tolerancia de simplificación especificada.

SVL_SPATIAL_SIMPLIFY es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_SPATIAL_SIMPLIFY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| consulta | entero | El ID de la consulta (comando COPY) que generó esta fila. |
| line_number | entero | Cuando se especifica la opción COPY `SIMPLIFY AUTO`, este valor es el número de registro del registro simplificado en el shapefile. |
| maximum_tolerance | double | El valor de tolerancia de distancia especificado en el comando COPY. Este es el valor de tolerancia máxima mediante la opción `SIMPLIFY AUTO` o el valor de tolerancia fijo mediante la opción `SIMPLIFY`. |
| inicial_size | entero | El tamaño en bytes del valor de datos `GEOMETRY` antes de la simplificación. |
| simplified | char(1) | Cuando se especifica la opción COPY `SIMPLIFY AUTO`, `t` si la geometría se simplificó correctamente, o, de lo contrario, `f`. Es posible que la geometría no se simplifique correctamente si después de la simplificación con la tolerancia máxima dada su tamaño sigue siendo mayor que el tamaño máximo de la geometría. |
| final_size | entero | Cuando se especifica la opción COPY `SIMPLIFY AUTO`, este es el tamaño en bytes de la geometría después de la simplificación. |
| final_tolerance | double | |

Consulta de ejemplo

La siguiente consulta devuelve la lista de registros que COPY simplificó.

```
SELECT * FROM svl_spatial_simplify WHERE query = pg_last_copy_id();
 query | line_number | maximum_tolerance | initial_size | simplified | final_size |
 final_tolerance
-------+-------------+-------------------+--------------+------------+------------
+----------------------
    20 |     1184704 |                -1 |      1513736 | t          |    1008808 |
 1.276386653895e-05
    20 |     1664115 |                -1 |      1233456 | t          |    1023584 |
 6.11707814796635e-06
```

## SVL_SPECTRUM_SCAN_ERROR

Puede consultar la vista del sistema SVL_SPECTRUM_SCAN_ERROR para obtener información acerca de los errores de análisis de Redshift Spectrum.

SVL_SPECTRUM_SCAN_ERROR es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_EXTERNAL_QUERY_ERROR. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

Muestra un ejemplo de errores registrados. El valor predeterminado es 10 entradas por consulta.

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la consulta. |
| consulta | entero | ID de la consulta que generó esta fila. |
| ubicación | character (128) | Ubicación de los datos que se están consultando. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| rowid | character (128) | La ubicación del error en el archivo. Las partes de `rowid` se separan con `:` (dos puntos), y es posible agregar otras partes en el futuro.<br>`row_offset :row_group :row_id`<br>Un row_offset es el desplazamiento (en bytes) de la fila dentro del archivo, y se establece en `-1` para formatos de archivo no admitidos. Una tabla se divide en row_groups, y cada grupo tiene filas con distintos row_ids. |
| colname | character (128) | Nombre de la columna que ha devuelto la consulta. |
| original_value | character (128) | Valor original consultado. |
| modified_value | character (128) | Valor modificado devuelto en función de la opción de configuración de control de datos especificada en la consulta. |
| Desencadenador | character (128) | Opción de control de datos especificada en la consulta. |
| acción | character (128) | Acción asociada a la opción de control de datos especificada en la consulta. |
| action_value | character (128) | Valor del parámetro de acción asociado a la opción de control de datos especificada en la consulta. |
| error_code | entero | Código del resultado de la opción de control de datos especificada en la consulta. |

Consulta de ejemplo

La siguiente consulta devuelve la lista de filas para las que se han realizado operaciones de control de datos.

```
SELECT * FROM svl_spectrum_scan_error;
```

La consulta devuelve resultados similares a los siguientes.

```
  userid  query     location                                    rowid   colname
         original_value             modified_value       trigger          action
                 action_valueerror_code
    100   1574007   s3://spectrum-uddh/league/spi_global_rankings.0:0     league_name
         Barclays Premier League    Barclays Premier Lea UNSPECIFIED      TRUNCATE
                                    156
    100   1574007   s3://spectrum-uddh/league/spi_global_rankings.0:0     league_nspi
         34595                      32767                UNSPECIFIED
OVERFLOW_VALUE                                  199
    100   1574007   s3://spectrum-uddh/league/spi_global_rankings.0:1     league_nspi
         34151                      32767                UNSPECIFIED
OVERFLOW_VALUE                                  199
    100   1574007   s3://spectrum-uddh/league/spi_global_rankings.0:2     league_name
         Barclays Premier League    Barclays Premier Lea UNSPECIFIED      TRUNCATE
                                    156
    100   1574007   s3://spectrum-uddh/league/spi_global_rankings.0:2     league_nspi
         33223                      32767                UNSPECIFIED
OVERFLOW_VALUE                                  199
    100   1574007   s3://spectrum-uddh/league/spi_global_rankings.0:3     league_name
         Barclays Premier League    Barclays Premier Lea UNSPECIFIED      TRUNCATE
                                    156
    100   1574007   s3://spectrum-uddh/league/spi_global_rankings.0:3     league_nspi
         32808                      32767                UNSPECIFIED
OVERFLOW_VALUE                                  199
    100   1574007   s3://spectrum-uddh/league/spi_global_rankings.0:4     league_nspi
         32790                      32767                UNSPECIFIED
OVERFLOW_VALUE                                  199
    100   1574007   s3://spectrum-uddh/league/spi_global_rankings.0:5     league_name
         Spanish Primera Division   Spanish Primera Divi UNSPECIFIED      TRUNCATE
                                    156
```

```
  100   1574007   s3://spectrum-uddh/league/spi_global_rankings.0:6     league_name
      Spanish Primera Division   Spanish Primera Divi UNSPECIFIED      TRUNCATE
                                 156
```

## SVL_STATEMENTTEXT

Utilice la vista SVL_STATEMENTTEXT para obtener un registro completo de todos los comandos SQL que se ejecutaron en el sistema.

La vista SVL_STATEMENTTEXT tiene la unión de todas las filas en las tablas STL_DDLTEXT, STL_QUERYTEXT y STL_UTILITYTEXT. Esta vista también incluye una combinación con la tabla STL_QUERY.

SVL_STATEMENTTEXT es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_QUERY_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | ID del usuario que generó la entrada. |
| xid | bigint | ID de la transacción asociado a la instrucción. |
| pid | entero | ID del proceso para la instrucción. |
| etiqueta | character(320) | Ya sea el nombre del archivo utilizado para ejecutar la consulta o una etiqueta definida con un comando SET QUERY_GROUP. Si la consulta no se basa en archivos o si no se establece el parámetro QUERY_GROUP, el valor del campo queda en blanco. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| starttime | marca de tiempo | Hora exacta en que la instrucción comenzó a ejecutarse, con 6 dígitos de precisión para las fracciones de segundos. Por ejemplo: **2009-06-12 11:29:19.131358** . |
| endtime | marca de tiempo | Hora exacta en que la instrucción finalizó la ejecución, con 6 dígitos de precisión para las fracciones de segundos. Por ejemplo: **2009-06-12 11:29:19.193640** . |
| sequence | entero | Cuando una única instrucción tiene más de 200 caractere s, se registran filas adicionales para esa instrucción. Secuencia 0 es la primera fila, 1 es la segunda y, así, sucesivamente. |
| type | varchar (10) | Tipo de instrucción SQL: **QUERY**, **DDL** o **UTILITY**. |
| texto | character(200) | Texto SQL, en incrementos de 200 caracteres. El campo puede contener caracteres especiales como una barra inversa(\\) y una nueva línea (\n). |

Consulta de ejemplo

En la siguiente consulta, se devuelven instrucciones DDL que se ejecutaron el 16 de junio de 2009:

```
select starttime, type, rtrim(text) from svl_statementtext
where starttime like '2009-06-16%' and type='DDL' order by starttime asc;

starttime                  | type |              rtrim
---------------------------|------|-------------------------------
2009-06-16 10:36:50.625097 | DDL  | create table ddltest(c1 int);
2009-06-16 15:02:16.006341 | DDL  | drop view alltickitjoin;
2009-06-16 15:02:23.65285  | DDL  | drop table sales;
2009-06-16 15:02:24.548928 | DDL  | drop table listing;
2009-06-16 15:02:25.536655 | DDL  | drop table event;
...
```

### Reconstruyendo el SQL almacenado

Para reconstruir el SQL almacenado en la columna `text` de SVL_STATEMENTTEXT, ejecute una instrucción SELECT para crear SQL desde 1 o más partes en la columna `text`. Antes de ejecutar el SQL reconstruido, reemplace cualquier caracter especial (\n) con una nueva línea. El resultado de la siguiente instrucción SELECT es filas de SQL reconstruido en el campo `query_statement`.

```
select LISTAGG(CASE WHEN LEN(RTRIM(text)) = 0 THEN text ELSE RTRIM(text) END, '')
 within group (order by sequence) AS query_statement
from SVL_STATEMENTTEXT where pid=pg_backend_pid();
```

Por ejemplo, la siguiente búsqueda selecciona 3 columnas. La búsqueda en si misma es más larga de 200 caracteres y se guarda en varias partes en VL_STATEMENTTEXT.

```
select
1 AS a012345678901234567890123456789012345678901234567890123456789012345678901234567890,
2 AS b012345678901234567890123456789012345678901234567890123456789012345678901234567890,
3 AS b012345678901234567890123456789012345678901234
FROM stl_querytext;
```

En este ejemplo, la búsqueda se guarda en 2 partes (filas) en la columna `text` de VL_STATEMENTTEXT.

```
select sequence, text from SVL_STATEMENTTEXT where pid = pg_backend_pid() order by
 starttime, sequence;
```

```
 sequence |
                text

----------
+-----------------------------------------------------------------------------------------------------------------------
        0 | select\n1 AS
 a012345678901234567890123456789012345678901234567890123456789012345678901234567890,\n2 AS
 b012345678901234567890123456789012345678901234567890123456789012345678901234567890,\n3 AS
 b012345678901234567890123456789012345678901234
        1 | \nFROM stl_querytext;
```

Para reconstruir el SQL almacenado en VL_STATEMENTTEXT, ejecute el siguiente SQL.

```
select LISTAGG(CASE WHEN LEN(RTRIM(text)) = 0 THEN text ELSE RTRIM(text) END, '')
 within group (order by sequence) AS text
from SVL_STATEMENTTEXT where pid=pg_backend_pid();
```

Para usar el SQL reconstruido resultante en su cliente, reemplace cualquier (\n) caracter especial con una nueva línea.

```
                    text

----------------------------------------------------------------------------------------------------
 select\n1 AS a01234567890123456789012345678901234567890123456789012345678901234567890,
\n2 AS b01234567890123456789012345678901234567890123456789012345678901234567890,\n3 AS
 b012345678901234567890123456789012345678901234\nFROM stl_querytext;
```

## SVL_STORED_PROC_CALL

Puede consultar la vista de sistema SVL_STORED_PROC_CALL para obtener información acerca de las llamadas de procedimientos almacenados, incluidos la hora de inicio, la hora de finalización y si se canceló una llamada. Cada llamada de procedimiento almacenado recibe un ID de consulta.

SVL_STORED_PROC_CALL es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_PROCEDURE_CALL. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | El ID del usuario cuyos privilegios se utilizaron para ejecutar la instrucción. Si esta llamada se anidó dentro de |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | un procedimiento almacenado de SECURITY DEFINER, entonces este es el id de usuario del propietario de ese procedimiento almacenado. |
| session_userid | entero | El ID del usuario que creó la sesión y el invocador de la llamada del procedimiento almacenado de nivel superior. |
| consulta | entero | El ID de la consulta de la llamada del procedimiento. |
| etiqueta | character(320) | Ya sea el nombre del archivo utilizado para ejecutar la consulta o una etiqueta definida con un comando SET QUERY_GROUP. Si la consulta no se basa en archivos o si no se establece el parámetro QUERY_GROUP, el valor del campo es el predeterminado. |
| xid | bigint | El ID de la transacción. |
| pid | entero | El ID del proceso. Por lo general, todas las consultas en una sesión se ejecutan en el mismo proceso, por lo que este valor suele permanecer constante si ejecuta una serie de consultas en la misma sesión. Después de determinados eventos internos, Amazon Redshift podría reiniciar una sesión activa y asignar un nuevo valor de PID. Para obtener más información, consulte STL_RESTARTED_SESSIONS. |
| base de datos | character (32) | El nombre de la base de datos al que estaba conectado el usuario cuando se emitió la consulta. |
| querytxt | character (4000) | El texto real de la consulta de la llamada del procedimiento. |
| starttime | marca de tiempo | La hora en UTC a la que la consulta comenzó a ejecutarse, con seis dígitos de precisión para los segundos fraccionados; por ejemplo, **`2009-06-12 11:29:19.131358`**. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| endtime | marca de tiempo | La hora en UTC a la que la consulta terminó de ejecutarse, con seis dígitos de precisión para los segundos fracciona dos; por ejemplo, **2009-06-12 11:29:19.131358.** |
| aborted | entero | Si el sistema detuvo un procedimiento almacenado o el usuario lo canceló, esta columna tendrá el valor 1. Si la consulta se ejecutó hasta su finalización, esta columna tendrá el valor 0. |
| from_sp_c all | entero | Si la llamada del procedimiento la invocó otra llamada de procedimiento, esta columna contiene el ID de la consulta de la llamada externa. De lo contrario, el campo es NULL. |

Consulta de ejemplo

La siguiente consulta devuelve el tiempo transcurrido en orden descendente y el estado de finalización de las llamadas del procedimiento almacenado correspondientes al último día.

```
select query, datediff(seconds, starttime, endtime) as elapsed_time, aborted,
 trim(querytxt) as call from svl_stored_proc_call where starttime >= getdate() -
 interval '1 day' order by 2 desc;

  query | elapsed_time | aborted |                                              call
--------+--------------+---------
+-------------------------------------------------------------------------------------
   4166 |            7 |       0 | call search_batch_status(35,'succeeded');
   2433 |            3 |       0 | call test_batch (123456)
   1810 |            1 |       0 | call prod_benchmark (123456)
   1836 |            1 |       0 | call prod_testing (123456)
   1808 |            1 |       0 | call prod_portfolio ('N', 123456)
   1816 |            1 |       1 | call prod_portfolio ('Y', 123456)
```

## SVL_STORED_PROC_MESSAGES

Puede consultar la vista del sistema SVL_STORED_PROC_MESSAGES para obtener información sobre los mensajes de procedimientos almacenados. Los mensajes generados se registran incluso

si se cancela la llamada al procedimiento almacenado. Cada llamada de procedimiento almacenado recibe un ID de consulta. Para obtener más información acerca de cómo establecer el nivel mínimo de mensajes registrados, consulte stored_proc_log_min_messages.

SVL_STORED_PROC_MESSAGES es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_PROCEDURE_MESSAGES. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| userid | entero | El ID del usuario cuyos privilegios se utilizaron para ejecutar la instrucción. Si esta llamada se anidó dentro de un procedimiento almacenado de SECURITY DEFINER, entonces este es el id de usuario del propietario de ese procedimiento almacenado. |
| session_userid | entero | El ID del usuario que creó la sesión y el invocador de la llamada del procedimiento almacenado de nivel superior. |
| pid | entero | El ID del proceso. |
| xid | bigint | El ID de la transacción de la consulta de la llamada del procedimiento. |
| consulta | entero | El ID de la consulta de la llamada del procedimiento. |
| recordtime | marca de tiempo | La hora en UTC a la que se generó el mensaje. |
| loglevel | entero | El valor numérico del nivel de registro del mensaje generado. Valores posibles: 20 para LOG, 30 para INFO, 40 para NOTICE, 50 para WARNING y 60 para EXCEPTION |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| loglevel_text | character (10) | Nivel de registro que se corresponde con el valor numérico de loglevel. Valores posibles: LOG, INFO, NOTICE, WARNING y EXCEPTION. |
| mensaje | character(1024) | El texto del mensaje generado. |
| linenum | entero | El número de línea de la instrucción generada. |
| querytext | character(500) | El texto real de la consulta de la llamada del procedimiento. |
| etiqueta | character(320) | Ya sea el nombre del archivo utilizado para ejecutar la consulta o una etiqueta definida con un comando SET QUERY_GROUP. Si la consulta no se basa en archivos o si no se establece el parámetro QUERY_GROUP, el valor del campo es el predeterminado. |
| aborted | entero | Si el sistema detuvo un procedimiento almacenado o el usuario lo canceló, esta columna tendrá el valor 1. Si la consulta se ejecutó hasta su finalización, esta columna tendrá el valor 0. |
| message_xid | bigint | El ID de la transacción del mensaje enviado. |

Consulta de ejemplo

Las siguientes instrucciones SQL muestran cómo utilizar SVL_STORED_PROC_MESSAGES para revisar los mensajes generados.

```
-- Create and run a stored procedure
CREATE OR REPLACE PROCEDURE test_proc1(f1 int) AS
$$
BEGIN
    RAISE INFO 'Log Level: Input f1 is %',f1;
    RAISE NOTICE 'Notice Level: Input f1 is %',f1;
```

```
    EXECUTE 'select invalid';
    RAISE NOTICE 'Should not print this';

EXCEPTION WHEN OTHERS THEN
     raise exception 'EXCEPTION level: Exception Handling';
END;
$$ LANGUAGE plpgsql;

-- Call this stored procedure
CALL test_proc1(2);

-- Show raised messages with level higher than INFO
SELECT query, recordtime, loglevel, loglevel_text, trim(message) as message, aborted
 FROM svl_stored_proc_messages
  WHERE loglevel > 30 AND query = 193 ORDER BY recordtime;

 query |         recordtime         | loglevel | loglevel_text |               message
              | aborted
-------+----------------------------+----------+---------------
+-----------------------------------+---------
   193 | 2020-03-17 23:57:18.277196 |       40 | NOTICE        | Notice Level: Input f1
 is 2          |       1
   193 | 2020-03-17 23:57:18.277987 |       60 | EXCEPTION     | EXCEPTION level:
 Exception Handling |       1
(2 rows)

-- Show raised messages at EXCEPTION level
SELECT query, recordtime, loglevel, loglevel_text, trim(message) as message, aborted
 FROM svl_stored_proc_messages
  WHERE loglevel_text = 'EXCEPTION' AND query = 193 ORDER BY recordtime;

 query |         recordtime         | loglevel | loglevel_text |               message
              | aborted
-------+----------------------------+----------+---------------
+-----------------------------------+---------
   193 | 2020-03-17 23:57:18.277987 |       60 | EXCEPTION     | EXCEPTION level:
 Exception Handling |       1
```

Las siguientes instrucciones SQL muestran cómo utilizar SVL_STORED_PROC_MESSAGES para revisar los mensajes generados con la opción SET cuando se crea un procedimiento almacenado. Como test_proc() tiene el nivel de registro mínimo de NOTICE, solo los mensajes de nivel NOTICE, WARNING y EXCEPTION se registran en SVL_STORED_PROC_MESSAGES.

```
-- Create a stored procedure with minimum log level of NOTICE
CREATE OR REPLACE PROCEDURE test_proc() AS
$$
BEGIN
    RAISE LOG 'Raise LOG messages';
    RAISE INFO 'Raise INFO messages';
    RAISE NOTICE 'Raise NOTICE messages';
    RAISE WARNING 'Raise WARNING messages';
    RAISE EXCEPTION 'Raise EXCEPTION messages';
    RAISE WARNING 'Raise WARNING messages again'; -- not reachable
END;
$$ LANGUAGE plpgsql SET stored_proc_log_min_messages = NOTICE;

-- Call this stored procedure
CALL test_proc();

-- Show the raised messages
SELECT query, recordtime, loglevel_text, trim(message) as message, aborted FROM
 svl_stored_proc_messages
  WHERE query = 149 ORDER BY recordtime;

 query |         recordtime         | loglevel_text |          message         |
 aborted
-------+----------------------------+---------------+--------------------------
+---------
   149 | 2020-03-16 21:51:54.847627 | NOTICE        | Raise NOTICE messages    |
 1
   149 | 2020-03-16 21:51:54.84766  | WARNING       | Raise WARNING messages   |
 1
   149 | 2020-03-16 21:51:54.847668 | EXCEPTION     | Raise EXCEPTION messages |
 1
(3 rows)
```

## SVL_TERMINATE

Registra la hora a la que un usuario cancela o termina un proceso.

SELECT PG_TERMINATE_BACKEND(pid), SELECT PG_CANCEL_BACKEND(pid) y CANCEL pid crea una entrada de registro en SVL_TERMINATE.

SVL_TERMINATE es visible solo para los superusuarios. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_QUERY_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| pid | entero | El ID del proceso cancelado o terminado. |
| eventtime | marca de tiempo | La hora a la que se cancela o termina el proceso. |
| userid | entero | El ID de usuario del usuario que ejecuta el comando. |
| type | cadena | El tipo de terminación. Puede ser CANCEL o TERMINATE. |

El siguiente comando muestra la última consulta cancelada.

```
select * from svl_terminate order by eventtime desc limit 1;
 pid  |         eventtime          | userid |  type
------+----------------------------+--------+--------
 8324 | 2020-03-24 09:42:07.298937 |      1 | CANCEL
(1 row)
```

## SVL_UDF_LOG

Registra los mensajes de error y advertencia definidos por el sistema que se generan durante la ejecución de la función definida por el usuario (UDF).

SVL_UDF_LOG es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_UDF_LOG. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| consulta | bigint | El ID de la consulta. Puede usar este ID para combinar distintas vistas y tablas del sistema. |
| mensaje | char (4096) | El mensaje generado por la función. |
| created | marca de tiempo | La hora en que se creó el registro. |
| traceback | char (4096) | Si está disponible, este valor brinda una pista de stack para la UDF. Para obtener más información, visite traceback en la biblioteca estándar de Python. |
| funcname | character (256) | El nombre de la UDF en ejecución. |
| node | entero | El nodo donde se generó el mensaje. |
| slice | entero | El sector donde se generó el mensaje. |
| seq | entero | La secuencia del mensaje en el sector. |

Consultas de ejemplo

En el siguiente ejemplo, se muestra cómo las UDF administran los errores definidos por el sistema. El primer bloque muestra la definición para una función UDF que devuelve la inversa de un argumento. Cuando ejecuta la función y proporciona un argumento 0, como se muestra en el

segundo bloque, la función devuelve un error. La tercera instrucción lee el mensaje de error que se registra en SVL_UDF_LOG

```
-- Create a function to find the inverse of a number

CREATE OR REPLACE FUNCTION  f_udf_inv(a int)
  RETURNS float IMMUTABLE
AS $$
   return 1/a
$$ LANGUAGE plpythonu;

-- Run the function with a 0 argument to create an error
Select f_udf_inv(0) from sales;

-- Query SVL_UDF_LOG to view the message

Select query, created, message::varchar
from svl_udf_log;

 query |          created           | message
-------+----------------------------
+------------------------------------------------------------
  2211 | 2015-08-22 00:11:12.04819  | ZeroDivisionError: long division or modulo by
 zero\nNone
```

En el siguiente ejemplo, se agrega el registro y un mensaje de advertencia a la UDF para que una operación de división en cero dé como resultado un mensaje de advertencia, en lugar de parar con un mensaje de error.

```
-- Create a function to find the inverse of a number and log a warning

CREATE OR REPLACE FUNCTION f_udf_inv_log(a int)
  RETURNS float IMMUTABLE
 AS $$
  import logging
  logger = logging.getLogger() #get root logger
  if a==0:
    logger.warning('You attempted to divide by zero.\nReturning zero instead of error.
\n')
    return 0
  else:
     return 1/a
```

```
$$ LANGUAGE plpythonu;
```

En el siguiente ejemplo, se ejecuta la función y, luego se consulta SVL_UDF_LOG para ver el mensaje.

```
-- Run the function with a 0 argument to trigger the warning
Select f_udf_inv_log(0) from sales;

-- Query SVL_UDF_LOG to view the message

Select query, created, message::varchar
from svl_udf_log;

query |          created           | message
------+----------------------------+---------------------------------
    0 | 2015-08-22 00:11:12.04819  | You attempted to divide by zero.
                                     Returning zero instead of error.
```

## SVL_USER_INFO

Puede recuperar datos sobre los usuarios de la base de datos de Amazon Redshift con la vista SVL_USER_INFO.

SVL_USER_INFO solo es visible para los superusuarios. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| usename | texto | Nombre de usuario de la función. |
| usesysid | entero | ID del usuario. |
| usecreatedb | booleano | Valor que indica si el usuario tiene permisos para crear bases de datos. |
| usesuper | booleano | Valor que indica si el usuario es un superusuario. |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| usecatupd | booleano | Valor que indica si el usuario puede actualizar los catálogos de sistema. |
| useconnlimit | texto | Número de conexiones que el usuario puede abrir. |
| syslogaccess | texto | Valor que indica si el usuario tiene acceso a los registros del sistema. Los dos valores posibles son `RESTRICTED` y `UNRESTRICTED`. `RESTRICTED` significa que los usuarios que no son superusuarios pueden ver sus propios registros. `UNRESTRICTED` significa que los usuarios que no son superusuarios pueden ver todos los registros de las vistas y tablas del sistema en las que tienen privilegios `SELECT`. |
| last_ddl_ts | marca de tiempo | Marca temporal de la instrucción de creación del lenguaje de definición de datos (DDL) ejecutada por el usuario. |
| sessiontimeout | entero | El tiempo máximo en segundos durante el cual una sesión permanece inactiva o en reposo antes de que se cierre. El 0 indica que no se establece tiempo de espera. Para obtener información acerca de la configuración de tiempo de espera en reposo o inactivo del clúster, consulte Cuotas y límites de Amazon Redshift en la Guía de administración de Amazon Redshift. |
| external_id | texto | Identificador único del usuario en el proveedor de identidades de terceros. |

Consultas de ejemplo

El siguiente comando recupera la información del usuario de SVL_USER_INFO.

```
SELECT * FROM SVL_USER_INFO;
```

# SVL_VACUUM_PERCENTAGE

La vista SVL_VACUUM_PERCENTAGE informa el porcentaje de bloques de datos asignados a una tabla después de realizar una limpieza. El número de porcentaje muestra cuánto espacio en disco se reclamó. Consulte el comando VACUUM para obtener más información acerca de la utilidad de limpieza.

Solo los superusuarios pueden ver SVL_VACUUM_PERCENTAGE. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

Algunos o todos los datos de esta tabla también están en la vista de monitoreo SYS SYS_VACUUM_HISTORY. Los datos de la vista de monitoreo SYS están formateados para que sean más fáciles de usar y entender. Se recomienda utilizar la vista de monitoreo SYS para las consultas.

Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| xid | bigint | ID de transacción para la instrucción de limpieza. |
| table_id | entero | ID de la tabla para la tabla que se limpió. |
| percentage | bigint | Porcentaje de bloques de datos después de una operación de limpieza (relacionados a la cantidad de bloqueos en una tabla antes de que se ejecute la operación). |

Consulta de ejemplo

La siguiente consulta muestra el porcentaje para una operación específica sobre la tabla 100238:

```
select * from svl_vacuum_percentage
where table_id=100238 and xid=2200;

xid  | table_id | percentage
-----+----------+------------
1337 |   100238 |         60
```

```
(1 row)
```

Luego de la operación de limpieza, la tabla contenía un 60% de los bloques originales.

# Tablas de catálogos de sistema

Temas


- PG_ATTRIBUTE_INFO
- PG_CLASS_INFO
- PG_DATABASE_INFO
- PG_DEFAULT_ACL
- PG_EXTERNAL_SCHEMA
- PG_LIBRARY
- PG_PROC_INFO
- PG_STATISTIC_INDICATOR
- PG_TABLE_DEF
- PG_USER_INFO
- Consulta de las tablas de catálogos


Los catálogos de sistema almacenan metadatos de esquema, como información acerca de tablas y columnas. Las tablas de catálogos de sistemas tienen un prefijo PG.

Los usuarios de Amazon Redshift pueden acceder a las tablas de catálogos de PostgreSQL estándar. Para obtener más información acerca de los catálogos de sistema de PostgreSQL, consulte PostgreSQL System Tables (Tablas de sistema de PostgreSQL)

## PG_ATTRIBUTE_INFO

PG_ATTRIBUTE_INFO es una vista de sistema de Amazon Redshift integrada a la tabla de catálogos de PostgreSQL PG_ATTRIBUTE y a la tabla de catálogos internos PG_ATTRIBUTE_ACL. PG_ATTRIBUTE_INFO incluye detalles sobre las columnas de una tabla o vista, incluidas las listas de control de acceso a columnas, si las hay.

### Columnas de la tabla

PG_ATTRIBUTE_INFO muestra la siguiente columna además de las columnas de PG_ATTRIBUTE.

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| attacl | aclitem[] | Los privilegios de acceso en el nivel de columna, si los hay, que se han concedido específicamente en esta columna. |

# PG_CLASS_INFO

PG_CLASS_INFO es una vista de sistema de Amazon Redshift integrada a las tablas de catálogos de PostgreSQL PG_CLASS y PG_CLASS_EXTENDED. PG_CLASS_INFO incluye detalles sobre la hora de creación y el estilo de distribución actual de la tabla. Para obtener más información, consulte Distribución de datos para la optimización de consultas.

Todos los usuarios pueden ver PG_CLASS_INFO. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

PG_CLASS_INFO muestra las siguientes columnas además de las columnas de PG_CLASS. La columna `oid` en PG_CLASS se llama `reloid` en la tabla PG_CLASS_INFO.

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| relcreationtime | marca de tiempo | Hora en UTC en que se creó la tabla. |
| releffectivediststyle | entero | El estilo de distribución de una tabla o, si la tabla usa distribución automática, el estilo de distribución actual asignado por Amazon Redshift. |

La columna RELEFFECTIVEDISTSTYLE en PG_CLASS_INFO indica el estilo de distribución actual de la tabla. Si la tabla usa una distribución automática, RELEFFECTIVEDISTSTYLE es 10, 11 o 12, lo que indica que el estilo de distribución efectivo es AUTO (ALL) o AUTO (EVEN) o AUTO (KEY). Si la tabla usa distribución automática, el estilo de distribución podría inicialmente mostrar AUTO (ALL),

para a continuación mostrar AUTO (EVEN) o AUTO (KEY) si se detecta que una columna puede resultar útil como clave de distribución.

En la siguiente tabla, se proporciona el estilo de distribución para cada valor de la tabla RELEFFECTIVEDISTSTYLE:

| RELEFFECTIVEDISTSTYLE | Estilo de distribución actual |
|---|---|
| 0 | EVEN |
| 1 | KEY |
| 8 | ALL |
| 10 | AUTO (ALL) |
| 11 | AUTO (EVEN) |
| 12 | AUTO (KEY) |

## Ejemplo

La siguiente consulta devuelve el estilo de distribución actual de las tablas del catálogo.

```
select reloid as tableid,trim(nspname) as schemaname,trim(relname) as
 tablename,reldiststyle,releffectivediststyle,
CASE WHEN "reldiststyle" = 0 THEN 'EVEN'::text
     WHEN "reldiststyle" = 1 THEN 'KEY'::text
     WHEN "reldiststyle" = 8 THEN 'ALL'::text
     WHEN "releffectivediststyle" = 10 THEN 'AUTO(ALL)'::text
     WHEN "releffectivediststyle" = 11 THEN 'AUTO(EVEN)'::text
     WHEN "releffectivediststyle" = 12 THEN 'AUTO(KEY)'::text ELSE '<<UNKNOWN>>'::text
 END as diststyle,relcreationtime
from pg_class_info a left join pg_namespace b on a.relnamespace=b.oid;
```

```
 tableid | schemaname | tablename | reldiststyle | releffectivediststyle | diststyle  |
      relcreationtime
---------+------------+-----------+--------------+-----------------------+------------
+---------------------------
```

```
 3638033 | public     | customer  |           0 |                    0 | EVEN       |
 2019-06-13 15:02:50.666718
 3638037 | public     | sales     |           1 |                    1 | KEY        |
 2019-06-13 15:03:29.595007
 3638035 | public     | lineitem  |           8 |                    8 | ALL        |
 2019-06-13 15:03:01.378538
 3638039 | public     | product   |           9 |                   10 | AUTO(ALL)  |
 2019-06-13 15:03:42.691611
 3638041 | public     | shipping  |           9 |                   11 | AUTO(EVEN) |
 2019-06-13 15:03:53.69192
 3638043 | public     | support   |           9 |                   12 | AUTO(KEY)  |
 2019-06-13 15:03:59.120695
(6 rows)
```

# PG_DATABASE_INFO

PG_DATABASE_INFO es una vista de sistema de Amazon Redshift que extiende la tabla de catálogos de PostgreSQL PG_DATABASE.

Todos los usuarios pueden ver PG_DATABASE_INFO.

## Columnas de la tabla

PG_DATABASE_INFO contiene las siguientes columnas además de las columnas en PG_DATABASE. La columna `oid` en PG_DATABASE se llama `datid` en la tabla PG_DATABASE_INFO. Para obtener más información, consulte la documentación de PostgreSQL.

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| datid | oid | El identificador de objeto (OID) usado internamente por las tablas del sistema. |
| datconnlimit | texto | El número máximo de conexiones simultáneas que pueden realizarse a esta base de datos. El valor 1 indica que no hay ningún límite. |

# PG_DEFAULT_ACL

Almacena información acerca de los privilegios de acceso predeterminados. Para obtener más información acerca de los privilegios de acceso predeterminados, consulte ALTER DEFAULT PRIVILEGES.

Todos los usuarios pueden ver PG_DEFAULT_ACL. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| defacluser | entero | El ID del usuario al que se aplican los privilegios mencionados. |
| defaclnamespace | oid | El ID de objeto del esquema en el que se aplican los privilegios predeterminados. Si no se especifica ningún esquema, el valor predeterminado es 0. |
| defaclobjtype | character | El tipo de objeto al que se aplican los privilegios predeterminados. Los valores válidos son los siguientes:<br>• r: relación (tabla o vista)<br>• f: función<br>• p: procedimiento almacenado |
| defaclacl | aclitem[] | Una cadena que define los privilegios predeterminados para el usuario o grupo de usuarios especificados y el tipo de objeto.<br>Si se conceden los privilegios a un usuario, la cadena es de la siguiente forma:<br>*{ username=privilegestring/grantor }*<br>username |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | El nombre del usuario al que se le conceden los privilegios. Si se omite el username (nombre de usuario), los privilegi os se conceden como PUBLIC.<br><br>Si se conceden los privilegios a un grupo de usuarios, la cadena es de la siguiente forma:<br><br>*{ "group groupname=privilegestring/g rantor" }*<br><br>privilegestring<br><br>Una cadena que especifica qué privilegios se conceden.<br><br>Los valores válidos son:<br><br>• a: INSERT (añadir)<br>• r: SELECT (leer)<br>• w: UPDATE (escribir)<br>• d: DELETE<br>• R: RULE<br>• x: concede el privilegio para crear una restricción de clave externa (REFERENCES).<br>• t: TRIGGER<br>• X: EXECUTE<br>• U: USAGE<br>• C: CREATE<br>• T: CREATE TEMP<br>• D: DROP<br>• P: TRUNCATE<br>• A: ALTER |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | • *: indica que el usuario que recibe el privilegio anterior puede, a su vez, conceder el mismo privilegio a otros (WITH GRANT OPTION).<br><br>Una cadena que contiene todos los caracteres del código de privilegio, ordenados por la posición de la máscara de bits, es algo parecido a arwdRxtXUCTDPA.<br><br>grantor<br><br>El nombre del usuario que concedió los privilegios.<br><br>En el siguiente ejemplo se indica que el usuario `admin` concedió todos los privilegios, incluido WITH GRANT OPTION, al usuario `dbuser`.<br><br>`dbuser=r*a*w*d*x*X*/admin` |

## Ejemplo

La siguiente consulta devuelve todos los privilegios predeterminados definidos para la base de datos.

```
select pg_get_userbyid(d.defacluser) as user,
n.nspname as schema,
case d.defaclobjtype when 'r' then 'tables' when 'f' then 'functions' end
as object_type,
array_to_string(d.defaclacl, ' + ')  as default_privileges
from pg_catalog.pg_default_acl d
left join pg_catalog.pg_namespace n on n.oid = d.defaclnamespace;

 user  | schema | object_type |              default_privileges
-------+--------+-------------+-------------------------------------------------------
 admin | tickit | tables      | user1=r/admin + "group group1=a/admin" + user2=w/admin
```

El resultado en el ejemplo anterior muestra que para todas las tablas nuevas creadas por el usuario `admin` en el esquema `tickit`, `admin` concede los privilegios SELECT a `user1`, los privilegios INSERT a `group1` y los privilegios UPDATE a `user2`.

# PG_EXTERNAL_SCHEMA

Almacena información acerca de los esquemas externos.

PG_EXTERNAL_SCHEMA es visible para todos los usuarios. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver los metadatos a los que tienen acceso. Para obtener más información, consulte CREATE EXTERNAL SCHEMA.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| esoid | oid | ID del esquema externo. |
| eskind | entero | Tipo de esquema externo. |
| esdbname | texto | Nombre de base de datos externa. |
| esoptions | texto | Opciones de esquemas externos. |

## Ejemplo

En el siguiente ejemplo, se muestran detalles de esquemas externos.

```
select esoid, nspname as schemaname, nspowner, esdbname as external_db, esoptions
from pg_namespace a,pg_external_schema b where a.oid=b.esoid;

esoid  | schemaname      | nspowner | external_db | esoptions

-------+-----------------+----------+-------------
+-------------------------------------------------------------
100134 | spectrum_schema |      100 | spectrum_db |
 {"IAM_ROLE":"arn:aws:iam::123456789012:role/mySpectrumRole"}
100135 | spectrum        |      100 | spectrumdb  |
 {"IAM_ROLE":"arn:aws:iam::123456789012:role/mySpectrumRole"}
```

```
100149 | external        |      100 | external_db |
 {"IAM_ROLE":"arn:aws:iam::123456789012:role/mySpectrumRole"}
```

# PG_LIBRARY

Almacena información acerca de las bibliotecas definidas por el usuario.

Todos los usuarios pueden ver PG_LIBRARY. Los superusuarios pueden ver todas las filas; los usuarios normales solo pueden ver sus datos. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| name | nombre | Nombre de la biblioteca. |
| language_oid | oid | Reservado para el uso del sistema. |
| file_store_id | entero | Reservado para el uso del sistema. |
| owner | entero | ID de usuario del propietario de la biblioteca. |

## Ejemplo

En el siguiente ejemplo se devuelve la información de las bibliotecas instaladas por el usuario.

```
select * from pg_library;

name       | language_oid | file_store_id | owner
-----------+--------------+---------------+------
f_urlparse |       108254 |          2000 |   100
```

# PG_PROC_INFO

PG_PROC_INFO es una vista de sistema de Amazon Redshift integrada a la tabla de catálogos de PostgreSQL PG_PROC y a la tabla de catálogo internos PG_PROC_EXTENDED. PG_PROC_INFO incluye detalles sobre funciones y procedimientos almacenados, incluida información relacionada con argumentos de salida, si los hubiere.

## Columnas de la tabla

PG_PROC_INFO muestra las siguientes columnas además de las columnas de PG_PROC. La columna `oid` en PG_PROC se llama `prooid` en la tabla PG_PROC_INFO.

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| prooid | oid | El ID del objeto de la función o procedimiento almacenado. |
| prokind | “char” | Un valor que indica el tipo de funciones o procedimientos almacenados. El valor es “f” para funciones normales, “p” para procedimientos almacenados y “a” para funciones agregadas. |
| proargmodes | "char"[ ] | Una matriz con los modos de los argumentos de los procedimientos, codificada como “i” para argumentos IN, “o” para argumentos OUT y “b” para argumentos INOUT. Si todos los argumentos son IN, este campo es NULL. Los subíndices corresponden a posiciones en la matriz proallargtypes. |
| proallargtypes | oid[ ] | Una matriz con los tipos de datos de los argumentos del procedimiento Esta matriz incluye todos los tipos de argumentos (incluidos los argumentos OUT e INOUT) No obstante, si todos los argumentos son argumentos IN, este campo es NULL. La suscripción se basa en uno. Por el contrario, proargtypes se suscriben desde cero. |

Los proargnames del campo en PG_PROC_INFO contienen los nombres de todos los tipos de argumentos (incluidos OUT e INOUT), si los hubiere.

# PG_STATISTIC_INDICATOR

Almacena información acerca del número de filas insertadas o eliminadas desde la última vez que se ejecutó ANALYZE. La tabla PG_STATISTIC_INDICATOR se actualiza frecuentemente siguiendo las operaciones DML, por lo que las estadísticas son aproximadas.

Solo los superusuarios pueden ver PG_STATISTIC_INDICATOR. Para obtener más información, consulte Visibilidad de datos en las tablas y vistas de sistema.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| stairelid | oid | ID de la tabla |
| stairows | float | Cantidad total de filas en la tabla. |
| staiins | float | Número de filas insertadas desde el último ANALYZE. |
| staidels | float | Número de filas eliminadas o actualizadas desde el último ANALYZE. |

## Ejemplo

En el siguiente ejemplo se devuelve la información de los cambios de la tabla desde el último ANALYZE.

```
select * from pg_statistic_indicator;

stairelid | stairows | staiins | staidels
----------+----------+---------+---------
   108271 |       11 |       0 |        0
   108275 |      365 |       0 |        0
   108278 |     8798 |       0 |        0
   108280 |    91865 |       0 |   100632
   108267 |    89981 |   49990 |     9999
   108269 |      808 |     606 |      374
   108282 |   152220 |   76110 |   248566
```

# PG_TABLE_DEF

Almacena información acerca de las columnas de la tabla.

PG_TABLE_DEF solo devuelve información acerca de las tablas que son visibles para el usuario. Si PG_TABLE_DEF no devuelve los resultados esperados, verifique que el parámetro search_path se estableció correctamente para incluir los esquemas relevantes.

Puede utilizar SVV_TABLE_INFO para consultar información más exhaustiva acerca de una tabla, incluidos el sesgo de distribución de datos, el sesgo de distribución de claves, el tamaño de tabla y las estadísticas.

## Columnas de la tabla

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| schemaname | nombre | Nombre del esquema. |
| tablename | nombre | Nombre de la tabla. |
| column | nombre | Nombre de la columna. |
| type | texto | Tipo de datos de la columna. |
| encoding | character (32) | Codificación de la columna. |
| distkey | booleano | True si esta columna es la clave de distribución para la tabla. |
| sortkey | entero | Orden de la columna en la clave de ordenación. Si la tabla utiliza una clave de ordenación compuesta, todas las columnas que forman parte de la clave de ordenación tienen un valor positivo que indica la posición de la columna en la clave de ordenación. Si la tabla utiliza una clave de ordenación intercalada, cada una de las columnas que formen parte de la clave de ordenación tienen un |

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| | | valor que puede ser positivo o negativo, donde el valor absoluto indica la posición de la columna en la clave de ordenación. Si es 0, la columna no forma parte de la clave de ordenación. |
| notnull | booleano | True si la columna tiene una restricción NOT NULL. |

## Ejemplo

En el siguiente ejemplo se muestran las columnas con clave de ordenación compuesta para la tabla LINEORDER_COMPOUND.

```
select "column", type, encoding, distkey, sortkey, "notnull"
from pg_table_def
where tablename = 'lineorder_compound'
and sortkey <> 0;

column       | type    | encoding | distkey | sortkey | notnull
-------------+---------+----------+---------+---------+--------
lo_orderkey  | integer | delta32k | false   |       1 | true
lo_custkey   | integer | none     | false   |       2 | true
lo_partkey   | integer | none     | true    |       3 | true
lo_suppkey   | integer | delta32k | false   |       4 | true
lo_orderdate | integer | delta    | false   |       5 | true
(5 rows)
```

En el siguiente ejemplo se muestran las columnas con clave de ordenación intercalada para la tabla LINEORDER_INTERLEAVED.

```
select "column", type, encoding, distkey, sortkey, "notnull"
from pg_table_def
where tablename = 'lineorder_interleaved'
and sortkey <> 0;

column       | type    | encoding | distkey | sortkey | notnull
-------------+---------+----------+---------+---------+--------
lo_orderkey  | integer | delta32k | false   |      -1 | true
lo_custkey   | integer | none     | false   |       2 | true
```

```
lo_partkey   | integer | none     | true    |      -3 | true
lo_suppkey   | integer | delta32k | false   |       4 | true
lo_orderdate | integer | delta    | false   |      -5 | true
(5 rows)
```

PG_TABLE_DEF solo devolverá información de las tablas en los esquemas que están incluidos en la ruta de búsqueda. Para obtener más información, consulte search_path.

Por ejemplo, suponga que crea un esquema y una tabla nuevos, luego consulta PG_TABLE_DEF.

```
create schema demo;
create table demo.demotable (one int);
select * from pg_table_def where tablename = 'demotable';

schemaname|tablename|column| type | encoding | distkey | sortkey | notnull
----------+---------+------+------+----------+---------+---------+--------
```

La consulta no devuelve filas para la nueva tabla. Analice la configuración de la `search_path`.

```
show search_path;

  search_path
---------------
 $user, public
(1 row)
```

Agregue el esquema demo a la ruta de búsqueda y ejecute la consulta de nuevo.

```
set search_path to '$user', 'public', 'demo';

select * from pg_table_def where tablename = 'demotable';

schemaname| tablename |column|  type   | encoding |distkey|sortkey| notnull
----------+-----------+------+---------+----------+-------+-------+--------
demo      | demotable | one  | integer | none     | f     |     0 | f
(1 row)
```

## PG_USER_INFO

PG_USER_INFO es una vista del sistema de Amazon Redshift que muestra información del usuario, como el ID de usuario y el tiempo de caducidad de la contraseña.

Solo los superusuarios pueden ver PG_USER_INFO.

### Columnas de la tabla

PG_USER_INFO contiene las siguientes columnas. Para obtener más información, consulte la documentación de PostgreSQL.

| Nombre de la columna | Tipo de datos: | Descripción |
|---|---|---|
| usename | nombre | El nombre de usuario. |
| usesysid | entero | El ID de usuario. |
| usecreatedb | booleano | True si el usuario puede crear bases de datos. |
| usesuper | booleano | True si el usuario es un superusuario. |
| usecatupd | booleano | True si el usuario puede actualizar los catálogos del sistema. |
| passwd | texto | La contraseña. |
| valuntil | abstime | Fecha y hora de caducidad de la contraseña. |
| useconfig | text[] | El valor predeterminado de la sesión para las variables en tiempo de ejecución. |
| useconnlimit | texto | Número de conexiones que el usuario puede abrir. |

## Consulta de las tablas de catálogos

Temas

- Ejemplos de consultas de catálogo

En general, puede combinar las tablas y las vistas de catálogos (relaciones cuyos nombres comienzan con **PG_**) con las tablas y las vistas de Amazon Redshift.

Las tablas de catálogos utilizan una serie de tipos de datos que Amazon Redshift no admite. Los siguientes tipos de datos se admiten cuando las consultas combinan las tablas de catálogos a las tablas de Amazon Redshift:

- bool
- "char"
- float4
- int2
- int4
- int8
- nombre
- oid
- texto
- varchar

Si escribe una consulta combinada que explícita o implícitamente hace referencia a una columna que tiene un tipo de datos no admitidos, la consulta devuelve un error. Las funciones SQL que se utilizan en algunas de las tablas de catálogos tampoco se admiten, a excepción de aquellas utilizadas por las tablas PG_SETTINGS y PG_LOCKS.

Por ejemplo, la tabla PG_STATS no puede consultarse en una combinación con una tabla de Amazon Redshift debido a las funciones no admitidas.

Las siguientes tablas y vistas de catálogos proporcionan información útil que se puede combinar a la información de las tablas de Amazon Redshift. Algunas de estas tablas permiten solo un acceso parcial debido a las restricciones de funciones y tipo de datos. Cuando consulte las tablas parcialmente accesibles, seleccione o consulte sus columnas con precaución.

Las siguientes tablas son completamente accesibles y no contienen tipos o funciones no admitidos:

- pg_attribute
- pg_cast
- pg_depend
- pg_description

- pg_locks
- pg_opclass

Las siguientes tablas son parcialmente accesibles y contienen algunos tipos, funciones y columnas de texto truncado no admitidos. Los valores en las columnas de textos se truncan en los valores varchar(256).

- pg_class
- pg_constraint
- pg_database
- pg_group
- pg_language
- pg_namespace
- pg_operator
- pg_proc
- pg_settings
- pg_statistic
- pg_tables
- pg_type
- pg_user
- pg_views

Las tablas de catálogos que no se mencionan aquí son inaccesibles o probablemente no sean de utilidad para los administradores de Amazon Redshift. Sin embargo, puede consultar cualquier tabla o vista de catálogos abiertamente si la consulta no incluye una combinación con una tabla de Amazon Redshift.

Puede utilizar las columnas OID en las tablas de catálogos Postgres como columnas de combinación. Por ejemplo, la condición de combinación `pg_database.oid = stv_tbl_perm.db_id` une el ID de objeto de la base de datos interna para cada fila de PG_DATABASE con la columna DB_ID visible en la tabla STV_TBL_PERM. Las columnas OID son claves primarias internas que no están visibles cuando selecciona desde la tabla. Las vistas de catálogo no tienen columnas OID.

Algunas funciones de Amazon Redshift deben ejecutarse solo en los nodos informáticos. Si una consulta hace referencia a una tabla creada por el usuario, el código SQL se ejecuta en los nodos de computación.

Una consulta que utiliza como referencia solamente tablas de catálogos (tablas con un prefijo PG, como por ejemplo PG_TABLE_DEF) o que no utiliza ninguna tabla se ejecuta únicamente en el nodo principal.

Si una consulta que utiliza una función de nodo informático no referencia una tabla definida por el usuario o una tabla del sistema de Amazon Redshift, devuelve el siguiente error.

```
[Amazon](500310) Invalid operation: One or more of the used functions must be applied
 on at least one user created table.
```

## Ejemplos de consultas de catálogo

En las siguientes consultas, se muestran algunas de las formas en las que puede consultar las tablas de catálogos para obtener información útil sobre una base de datos de Amazon Redshift.

Vista del ID de la tabla, el nombre de la base de datos, los nombre del esquema y de la tabla

La siguiente definición de vista combina la tabla de sistema STV_TBL_PERM con las tablas de catálogos de sistemas PG_CLASS, PG_NAMESPACE y PG_DATABASE para devolver el ID de la tabla, el nombre de la base de datos y los nombres del esquema y de la tabla.

```
create view tables_vw as
select distinct(stv_tbl_perm.id) table_id
,trim(pg_database.datname)   db_name
,trim(pg_namespace.nspname)   schema_name
,trim(pg_class.relname)   table_name
from stv_tbl_perm
join pg_class on pg_class.oid = stv_tbl_perm.id
join pg_namespace on pg_namespace.oid = pg_class.relnamespace
join pg_database on pg_database.oid = stv_tbl_perm.db_id;
```

En el siguiente ejemplo se devuelve la información para el ID de tabla 117855.

```
select * from tables_vw where table_id = 117855;
```

```
table_id | db_name   | schema_name | table_name
```

```
---------+-----------+-------------+-----------
  117855 |       dev | public      | customer
```

Cómo mostrar el número de columnas por tabla de Amazon Redshift

La siguiente consulta combina algunas tablas de catálogos para averiguar cuántas columnas contiene cada tabla de Amazon Redshift. Los nombres de tabla de Amazon Redshift se almacenan en PG_TABLES y STV_TBL_PERM; siempre que sea posible, utilice PG_TABLES para devolver nombres de tablas de Amazon Redshift.

Esta consulta no incluye tablas de Amazon Redshift.

```
select nspname, relname, max(attnum) as num_cols
from pg_attribute a, pg_namespace n, pg_class c
where n.oid = c.relnamespace and  a.attrelid = c.oid
and c.relname not like '%pkey'
and n.nspname not like 'pg%'
and n.nspname not like 'information%'
group by 1, 2
order by 1, 2;

nspname | relname  | num_cols
--------+----------+----------
public  | category |        4
public  | date     |        8
public  | event    |        6
public  | listing  |        8
public  | sales    |       10
public  | users    |       18
public  | venue    |        5
(7 rows)
```

Cómo mostrar los esquemas y las tablas en una base de datos

La siguiente consulta combina STV_TBL_PERM con algunas tablas PG para devolver una lista de tablas en la base de datos TICKIT y sus nombres de esquema (columna NSPNAME). La consulta también devuelve el número total de filas en cada tabla. (Esta consulta es útil cuando varios esquemas en su sistema tienen los mismos nombres de tablas).

```
select datname, nspname, relname, sum(rows) as rows
```

```
from pg_class, pg_namespace, pg_database, stv_tbl_perm
where pg_namespace.oid = relnamespace
and pg_class.oid = stv_tbl_perm.id
and pg_database.oid = stv_tbl_perm.db_id
and datname ='tickit'
group by datname, nspname, relname
order by datname, nspname, relname;

datname | nspname | relname  |  rows
--------+---------+----------+--------
tickit  | public  | category |     11
tickit  | public  | date     |    365
tickit  | public  | event    |   8798
tickit  | public  | listing  | 192497
tickit  | public  | sales    | 172456
tickit  | public  | users    |  49990
tickit  | public  | venue    |    202
(7 rows)
```

Cómo mostrar los ID de tablas, los tipos de datos y los nombres de las columnas y de las tablas

En la siguiente consulta se muestra información acerca de cada tabla de usuario y sus columnas: el ID de tabla, el nombre de la tabla, los nombres de sus columnas y el tipo de datos de cada columna:

```
select distinct attrelid, rtrim(name), attname, typname
from pg_attribute a, pg_type t, stv_tbl_perm p
where t.oid=a.atttypid and a.attrelid=p.id
and a.attrelid between 100100 and 110000
and typname not in('oid','xid','tid','cid')
order by a.attrelid asc, typname, attname;

attrelid |  rtrim   |    attname     |  typname
---------+----------+----------------+-----------
  100133 | users    | likebroadway   | bool
  100133 | users    | likeclassical  | bool
  100133 | users    | likeconcerts   | bool
...
  100137 | venue    | venuestate     | bpchar
  100137 | venue    | venueid        | int2
  100137 | venue    | venueseats     | int4
  100137 | venue    | venuecity      | varchar
...
```

Cómo contar el número de bloques de datos por cada columna en una tabla

La siguiente consulta combina la tabla STV_BLOCKLIST a PG_CLASS para devolver información de almacenamiento de las columnas en la tabla SALES.

```
select col, count(*)
from stv_blocklist s, pg_class p
where s.tbl=p.oid and relname='sales'
group by col
order by col;

col | count
----+-------
  0 |     4
  1 |     4
  2 |     4
  3 |     4
  4 |     4
  5 |     4
  6 |     4
  7 |     4
  8 |     4
  9 |     8
 10 |     4
 12 |     4
 13 |     8
(13 rows)
```

# Referencia de la configuración

Puede utilizar configuraciones para ayudar a personalizar el entorno. Con Amazon Redshift, puede personalizar y optimizar el entorno de almacenamiento de datos mediante la configuración de varios parámetros y ajustes. La referencia de configuración describe las propiedades del clúster, los parámetros de la base de datos y las opciones de configuración de la administración de la carga de trabajo (WLM) disponibles. Puede consultar esta referencia para afinar el rendimiento, la seguridad y la asignación de recursos en función de los requisitos específicos. La siguiente referencia proporciona una guía detallada sobre cómo modificar estas configuraciones para lograr la configuración de almacenamiento de datos deseada.

Temas


- Modificación de la configuración del servidor
- analyze_threshold_percent
- cast_super_null_on_error
- datashare_break_glass_session_var
- datestyle
- default_geometry_encoding
- describe_field_name_in_uppercase
- downcase_delimited_identifier
- enable_case_sensitive_identifier
- enable_case_sensitive_super_attribute
- enable_numeric_rounding
- enable_result_cache_for_session
- enable_vacuum_boost
- error_on_nondeterministic_update
- extra_float_digits
- interval_forbid_composite_literals
- json_serialization_enable
- json_serialization_parse_nested_strings
- max_concurrency_scaling_clusters
- max_cursor_result_set_size

- mv_enable_aqmv_for_session
- navegate_super_null_on_error
- parse_super_null_on_error
- pg_federation_repeatable_read
- query_group
- search_path
- spectrum_enable_pseudo_columns
- enable_spectrum_oid
- spectrum_query_maxerror
- statement_timeout
- stored_proc_log_min_messages
- timezone
- use_fips_ssl
- wlm_query_slot_count

## Modificación de la configuración del servidor

Puede cambiar la configuración del servidor de las siguientes formas:

- Mediante el uso de un comando SET para invalidar una configuración solamente durante el tiempo que dure su sesión actual.

  Por ejemplo:

  ```
  set extra_float_digits to 2;
  ```

- Mediante la modificación de la configuración del grupo de parámetros del clúster. La configuración del grupo de parámetros incluye parámetros adicionales que el usuario puede configurar. Para obtener más información, consulte Grupos de parámetros de Amazon Redshift en la Guía de administración de Amazon Redshift.
- Mediante el uso del comando ALTER USER para establecer un parámetro de configuración en un valor nuevo para todas las sesiones ejecutadas por el usuario especificado.

  ```
  ALTER USER username SET parameter { TO | = } { value | DEFAULT }
  ```

Utilice el comando SHOW para ver la configuración de parámetros actual. Utilice SHOW ALL para ver todos los ajustes que puede configurar con el comando SET.

```
SHOW ALL;

name                       | setting
---------------------------+--------------
analyze_threshold_percent | 10
datestyle                  | ISO, MDY
extra_float_digits         | 2
query_group                | default
search_path                | $user, public
statement_timeout          | 0
timezone                   | UTC
wlm_query_slot_count       | 1
```

> **Note**
>
> Tenga en cuenta que los parámetros de configuración se aplican a la base de datos a la que se conecta en el almacenamiento de datos.

# analyze_threshold_percent

## Valores (predeterminados en negrita)

10, de 0 a 100,0

## Descripción

Establece el umbral para el porcentaje de filas modificadas por el análisis de una tabla. Para reducir el tiempo de procesamiento y mejorar el rendimiento general del sistema, Amazon Redshift omite el análisis ANALYZE de cualquier tabla que tenga un porcentaje de filas modificadas inferior a lo especificado por `analyze_threshold_percent`. Por ejemplo, si una tabla contiene 100 000 000 filas y 9 000 000 tienen cambios desde que se ejecutó el último análisis ANALYZE, la tabla se omite de forma predeterminada ya que ha cambiado menos del 10 por ciento de las filas. Para analizar las tablas cuando solo han cambiado unas pocas filas, establezca `analyze_threshold_percent` en un número pequeño elegido arbitrariamente. Por ejemplo, si establece `analyze_threshold_percent` como 0,01, no se omitirá una tabla con 100 000 000 filas

si han cambiado al menos 10 000 filas. Para analizar todas las tablas, aunque no haya cambiado ninguna fila, establezca `analyze_threshold_percent` en 0.

Puede modificar el parámetro `analyze_threshold_percent` solo para la sesión actual usando un comando SET. El parámetro no se puede modificar en un grupo de parámetros.

### Ejemplo

```
set analyze_threshold_percent to 15;
set analyze_threshold_percent to 0.01;
set analyze_threshold_percent to 0;
```

## cast_super_null_on_error

### Valores (predeterminados en negrita)

on (encendido), off (apagado)

### Descripción

Especifica que, cuando intenta acceder a un miembro inexistente de un objeto o elemento de una matriz, Amazon Redshift devuelve un valor NULL si la consulta se ejecuta en el modo relajado predeterminado.

## datashare_break_glass_session_var

### Valores (predeterminados en negrita)

No hay valores predeterminados. El valor puede ser cualquier cadena de caracteres generada por Amazon Redshift cuando se lleva a cabo una operación que no se recomienda, como se describe a continuación.

### Descripción

Aplica un permiso que permite ciertas operaciones que, por lo general, no se recomiendan para un recurso compartido de datos de AWS Data Exchange.

En general, recomendamos que no elimine ni modifique un recurso compartido de datos de AWS Data Exchange mediante la instrucción DROP DATASHARE o ALTER DATASHARE SET

PUBLICACCESIBLE. Para permitir que se elimine o modifique un recurso compartido de datos de AWS Data Exchange y se desactive la configuración de acceso público, establezca la variable `datashare_break_glass_session_var` en un valor único. Amazon Redshift genera este valor único y se proporciona en un mensaje de error luego del intento inicial de la operación en cuestión.

Después de establecer la variable en el valor único generado, vuelva a ejecutar la instrucción DROP DATASHARE o ALTER DATASHARE.

Para obtener más información, consulte Notas de uso de ALTER DATASHARE o Notas de uso de DROP DATASHARE.

### Ejemplo

```
set datashare_break_glass_session_var to '620c871f890c49';
```

## datestyle

### Valores (predeterminados en negrita)

Especificación de formato (ISO, Postgres, SQL o German) y orden año/mes/día (DMY, MDY, YMD).

- ISO: usa el estilo de fecha AAAA-MM-DD HH:MM:SS.
- Postgres: usa el estilo de fecha MM-DD HH:MM:SS AAAA.
- SQL: usa el estilo de fecha MM-DD-AAAA HH:MM:SS.
- German: usa el estilo de fecha DD-MM-AAAA HH:MM:SS.

### Descripción

Configura el formato de visualización para los valores de fecha y hora, como también las reglas para interpretar valores ambiguos de entrada de las fechas. La cadena contiene dos parámetros que se pueden modificar juntos o por separado.

### Ejemplo

```
show datestyle;
DateStyle
-----------
```

```
ISO, MDY
(1 row)

set datestyle to 'SQL,DMY';
```

# default_geometry_encoding

## Valores (predeterminados en negrita)

1, 2

## Descripción

Es la configuración de una sesión que especifica si las geometrías espaciales creadas durante esta sesión se codifican con un cuadro delimitador. Si `default_geometry_encoding` está configurado en 1, las geometrías no se codifican con un cuadro delimitador. Si `default_geometry_encoding` está configurado en 2, las geometrías se codifican con un cuadro delimitador. Para obtener más información acerca de la compatibilidad con los cuadros delimitadores, consulte Cuadro delimitador.

# describe_field_name_in_uppercase

## Valores (predeterminados en negrita)

off (false), on (true)

## Descripción

Especifica si los nombres de columna devueltos por las instrucciones SELECT están en mayúsculas o minúsculas. Si el parámetro está en on (encendido), los nombres de columna se devuelven en mayúsculas. Si el parámetro está en off (apagado), los nombres de columna se devuelven en minúsculas. Amazon Redshift almacena los nombres de columna en minúsculas independientemente de la configuración de `describe_field_name_in_uppercase`.

## Ejemplo

```
set describe_field_name_in_uppercase to on;

show describe_field_name_in_uppercase;
```

```
DESCRIBE_FIELD_NAME_IN_UPPERCASE
--------------------------------
on
```

# downcase_delimited_identifier

## Valores (predeterminados en negrita)

on (encendido), off (apagado)

## Descripción

Esta configuración se está retirando. En su lugar, utilice `enable_case_sensitive_identifier`.

Permite que el superanalizador lea los campos JSON que están en mayúsculas o en mayúsculas y minúsculas. También habilita la compatibilidad con consultas federadas para bases de datos PostgreSQL compatibles con nombres en mayúsculas y minúsculas de bases de datos, esquemas, tablas y columnas. Para utilizar identificadores que distinguen entre mayúsculas y minúsculas, configure este parámetro en off (apagado).

## Notas de uso

- Si utiliza características de seguridad de nivel de fila o de enmascaramiento dinámico de datos, le recomendamos que establezca el valor `downcase_delimited_identifier` en el grupo de parámetros del clúster o grupo de trabajo. Esto garantiza que `downcase_delimited_identifier` se mantenga constante al crear y adjuntar una política y, a continuación, al consultar una relación a la que se haya aplicado una política. Para obtener información sobre seguridad de nivel de fila, consulte Seguridad de nivel básico. Para obtener información sobre el enmascaramiento dinámico de datos, consulte Enmascaramiento de datos dinámico.
- Cuando se desactiva `downcase_delimited_identifier` y se crea una tabla, se pueden establecer nombres de columnas que distingan mayúsculas de minúsculas. Cuando se activa `downcase_delimited_identifier` y se consulta la tabla, los nombres de las columnas aparecen en minúsculas. Esto puede producir resultados de consulta diferentes a los que se obtienen cuando `downcase_delimited_identifier` está desactivado. Considere el siguiente ejemplo:

```
SET downcase_delimited_identifier TO off;
```

```
--Amazon Redshift preserves case for column names and other identifiers.

--Create a table with two columns that are identical except for the case.
CREATE TABLE t ("c" int, "C" int);

INSERT INTO t VALUES (1, 2);

SELECT * FROM t;

 c | C
---+---
 1 | 2
(1 row)

SET enable_downcase_delimited_identifier TO on;
--Amazon Redshift no longer preserves case for column names and other identifiers.

SELECT * FROM t;

 c | c
---+---
 1 | 1
(1 row)
```

- Recomendamos que los usuarios habituales que consulten tablas con políticas de enmascaramiento dinámico de datos o de seguridad en el nivel de fila tengan la configuración downcase_delimited_identifier predeterminada. Para obtener más información sobre seguridad en el nivel de fila, consulte Seguridad de nivel básico. Para obtener información sobre el enmascaramiento dinámico de datos, consulte Enmascaramiento de datos dinámico.

# enable_case_sensitive_identifier

## Valores (predeterminados en negrita)

true, false

## Descripción

Valor de configuración que determina si los identificadores de nombres de bases de datos, esquemas, tablas y columnas distinguen entre mayúsculas y minúsculas. Las letras mayúsculas y minúsculas de los identificadores de nombres se conservan cuando los identificadores se encierran

entre comillas dobles y `enable_case_sensitive_identifier` se establece en `true`. Las mayúsculas y minúsculas de los identificadores de nombres no se conservan y, en cambio, se convierten a minúsculas cuando los identificadores no se encierran entre comillas dobles o cuando `enable_case_sensitive_identifier` se establece en `false`.

El caso de un nombre de usuario entre comillas dobles se conserva siempre, independientemente del valor de la opción de configuración `enable_case_sensitive_identifier`.

## Ejemplos

En el siguiente ejemplo, se muestra cómo crear y utilizar identificadores que distinguen entre mayúsculas y minúsculas para el nombre de una columna y una tabla.

```
-- To create and use case sensitive identifiers
SET enable_case_sensitive_identifier TO true;

-- Create tables and columns with case sensitive identifiers
CREATE TABLE public."MixedCasedTable" ("MixedCasedColumn" int);

INSERT INTO public."MixedCasedTable" VALUES (1);
INSERT INTO public."MixedCasedTable" VALUES (2);
INSERT INTO public."MixedCasedTable" VALUES (3);
INSERT INTO public."MixedCasedTable" VALUES (4);
INSERT INTO public."MixedCasedTable" VALUES (5);

-- Now query with case sensitive identifiers
SELECT "MixedCasedColumn" FROM public."MixedCasedTable";

MixedCasedColumn
------------------
1
2
3
4
5

(5 rows)


SELECT * FROM public."MixedCasedTable" WHERE "MixedCasedColumn" = 1;

mixedcasedcolumn
------------------
```

```
1

(1 row)
```

En el siguiente ejemplo, se muestra lo que sucede cuando no se conservan las letras mayúsculas y minúsculas de los identificadores.

```
-- To not use case sensitive identifiers
RESET enable_case_sensitive_identifier;

-- Mixed case identifiers are lowercased despite double quotation marks
CREATE TABLE "MixedCasedTable2" ("MixedCasedColumn" int);

CREATE TABLE MixedCasedTable2 (MixedCasedColumn int);

ERROR:  Relation "mixedcasedtable2" already exists


SELECT "MixedCasedColumn" FROM "MixedCasedTable2";

 mixedcasedcolumn
------------------
(0 rows)


SELECT MixedCasedColumn FROM MixedCasedTable2;

 mixedcasedcolumn
------------------
(0 rows)
```

## Notas de uso

- Si utiliza la actualización automática para las vistas materializadas, le recomendamos que establezca el valor `enable_case_sensitive_identifier` en el grupo de parámetros del clúster o grupo de trabajo. De este modo, se garantiza que `enable_case_sensitive_identifier` permanezca constante cuando se actualicen las vistas materializadas. Para obtener información sobre la actualización automática de las vistas materializadas, consulte Actualización de una vista materializada. Para obtener información sobre el establecimiento de los valores de configuración, consulte Grupos de parámetros de Amazon Redshift en la Guía de administración de Amazon Redshift.

- Si utiliza características de seguridad de nivel de fila o de enmascaramiento dinámico de datos, le recomendamos que establezca el valor `enable_case_sensitive_identifier` en el grupo de parámetros del clúster o grupo de trabajo. Esto garantiza que `enable_case_sensitive_identifier` se mantenga constante al crear y adjuntar una política y, a continuación, al consultar una relación a la que se haya aplicado una política. Para obtener información sobre seguridad de nivel de fila, consulte Seguridad de nivel básico. Para obtener información sobre el enmascaramiento dinámico de datos, consulte Enmascaramiento de datos dinámico.
- Cuando se activa `enable_case_sensitive_identifier` y se crea una tabla, se pueden establecer nombres de columnas que distingan mayúsculas de minúsculas. Cuando se desactiva `enable_case_sensitive_identifier` y se consulta la tabla, los nombres de las columnas aparecen en minúsculas. Esto puede producir resultados de consulta diferentes a los que se obtienen cuando `enable_case_sensitive_identifier` está activado. Considere el siguiente ejemplo:

```
SET enable_case_sensitive_identifier TO on;
--Amazon Redshift preserves case for column names and other identifiers.

--Create a table with two columns that are identical except for the case.
CREATE TABLE t ("c" int, "C" int);

INSERT INTO t VALUES (1, 2);

SELECT * FROM t;

 c | C
---+---
 1 | 2
(1 row)

SET enable_case_sensitive_identifier TO off;
--Amazon Redshift no longer preserves case for column names and other identifiers.

SELECT * FROM t;

 c | c
---+---
 1 | 1
(1 row)
```

- Recomendamos que los usuarios habituales que consulten tablas con políticas de enmascaramiento dinámico de datos o de seguridad en el nivel de fila tengan la configuración enable_case_sensitive_identifier predeterminada. Para obtener información sobre seguridad de nivel de fila, consulte Seguridad de nivel básico. Para obtener información sobre el enmascaramiento dinámico de datos, consulte Enmascaramiento de datos dinámico.
- Para hacer referencia a los identificadores con mayúsculas y minúsculas combinadas mediante la notación de puntos, escriba cada identificador con distinción entre mayúsculas y minúsculas entre comillas dobles. Por ejemplo, `public."MixedCasedTable"."MixedCasedColumn"`.

# enable_case_sensitive_super_attribute

## Valores (predeterminados en negrita)

true, false

## Descripción

Un valor de configuración que determina si en la navegación por las estructuras de tipos de datos SUPER con nombres de atributo sin delimitar se distingue entre mayúsculas y minúsculas. Al establecer `enable_case_sensitive_super_attribute` a `true`, en la navegación por las estructuras de tipos SUPER con nombres de atributo sin delimitar se distingue entre mayúsculas y minúsculas. Al establecer el valor a `false`, en la navegación por las estructuras de tipos SUPER con nombres de atributo sin delimitar no se distingue entre mayúsculas y minúsculas.

Cuando se incluye un nombre de atributo entre comillas dobles y se establece `enable_case_sensitive_identifier` a `true`, siempre se conservan las mayúsculas y minúsculas, independientemente del establecimiento de la opción de configuración `enable_case_sensitive_super_attribute`.

`enable_case_sensitive_super_attribute` solo se aplica a columnas con el tipo datos SUPER. Para el resto de columnas, considere utilizar `enable_case_sensitive_identifier`.

Para obtener más información acerca del tipo de datos SUPER, consulte Tipo SUPER y Datos semiestructurados en Amazon Redshift.

## Ejemplos

En el siguiente ejemplo, se muestran los resultados de seleccionar valore SUPER con `enable_case_sensitive_super_attribute` habilitado y con ese atributo deshabilitado.

```
--Create a table with a SUPER column.
CREATE TABLE tbl (col SUPER);

--Insert values.
INSERT INTO tbl VALUES (json_parse('{
 "A": "HELLO", "a": "123"
}'));

SET enable_case_sensitive_super_attribute TO ON;

SELECT col.A FROM tbl;
  a
-----
 "HELLO"
(1 row)

SELECT col.a FROM tbl;
  a
-----
 "123"
(1 row)

SET enable_case_sensitive_super_attribute TO OFF;

SELECT col.A FROM tbl;
  a
-----
 "123"
(1 row)

SELECT col.a FROM tbl;
  a
-----
 "123"
(1 row)
```

## Notas de uso

- Las vistas y las vistas materializadas siguen el valor de `enable_case_sensitive_super_attribute` en el momento de crearlas. Las vistas de enlace en tiempo de ejecución, los procedimientos almacenados y las funciones definidas por el usuario siguen el valor de `enable_case_sensitive_super_attribute` en el momento de la consulta.

- Si utiliza la actualización automática para las vistas materializadas, le recomendamos que establezca el `enable_case_sensitive_identifier value` en el grupo de parámetros del clúster o grupo de trabajo. De este modo, se garantiza que `enable_case_sensitive_identifier` permanezca constante cuando se actualicen las vistas materializadas. Para obtener información sobre la actualización automática de las vistas materializadas, consulte Actualización de una vista materializada. Para obtener información sobre el establecimiento de los valores de configuración, consulte Grupos de parámetros de Amazon Redshift en la Guía de administración de Amazon Redshift.
- El nombre de la columna en los resultados de la instrucción siempre se cambia a minúsculas, independientemente del valor de `enable_case_sensitive_super_attribute`. Para que en el nombre de la columna también se distinga entre mayúsculas y minúsculas, active `enable_case_sensitive_identifier`.
- Recomendamos que los usuarios habituales que consulten tablas con políticas de seguridad en el nivel de fila asociadas tengan la configuración `enable_case_sensitive_identifier` predeterminada. Para obtener más información sobre seguridad en el nivel de fila, consulte Seguridad de nivel básico.

# enable_numeric_rounding

## Valores (predeterminados en negrita)

on (verdadero), off (falso)

## Descripción

Especifica si se debe utilizar el redondeo numérico. Si `enable_numeric_rounding` es on, Amazon Redshift redondea los valores NUMERIC al convertirlos a otros tipos numéricos, como INTEGER o DECIMAL. Si `enable_numeric_rounding` es `off`, Amazon Redshift trunca los valores NUMERIC al convertirlos a otros tipos numéricos. Para obtener más información sobre tipos numéricos, consulte Tipos numéricos.

## Ejemplo

```
--Create a table and insert the numeric value 1.5 into it.
CREATE TABLE t (a numeric(10, 2));

INSERT INTO t VALUES (1.5);
```

```
SET enable_numeric_rounding to ON;
--Amazon Redshift now rounds NUMERIC values when casting to other numeric types.

SELECT a::int FROM t;

 a
---
 2
(1 row)


SELECT a::decimal(10, 0) FROM t;

 a
---
 2
(1 row)


 SELECT a::decimal(10, 5) FROM t;

    a
---------
 1.50000
(1 row)


SET enable_numeric_rounding to OFF;
--Amazon Redshift now truncates NUMERIC values when casting to other numeric types.

SELECT a::int FROM t;

 a
---
 1
(1 row)


SELECT a::decimal(10, 0) FROM t;

 a
---
 1
```

```
(1 row)


SELECT a::decimal(10, 5) FROM t;

    a
---------
 1.50000
(1 row)
```

# enable_result_cache_for_session

## Valores (predeterminados en negrita)

on (true), off (false)

## Descripción

Especifica si se va a utilizar el almacenamiento en caché de los resultados de las consultas. Si `enable_result_cache_for_session` está configurado en on, cuando se envía una consulta, Amazon Redshift busca en el almacenamiento en caché una copia válida de los resultados de la consulta. Si se encuentra alguna coincidencia en la caché de resultados, Amazon Redshift utiliza estos resultados almacenados en caché y no ejecuta la consulta. Si `enable_result_cache_for_session` está configurado en `off`, Amazon Redshift no tiene en cuenta el almacenamiento en caché de los resultados y ejecuta todas las consultas cuando se envían.

## Ejemplo

```
SET enable_result_cache_for_session TO off;
--Amazon Redshift now ignores the results cache
```

# enable_vacuum_boost

## Valores (predeterminados en negrita)

false, true

## Descripción

Especifica si se debe habilitar la opción de vacuum boost para todos los comandos VACUUM que se ejecutan en una sesión. Si `enable_vacuum_boost` está configurado en `true`, Amazon Redshift ejecuta todos los comandos VACUUM en la sesión con la opción BOOST. Si `enable_vacuum_boost` está configurado en `false`, Amazon Redshift no se ejecuta con la opción BOOST de forma predeterminada. Para obtener más información acerca de la opción BOOST, consulte VACUUM.

# error_on_nondeterministic_update

## Valores (predeterminados en negrita)

false, true

## Descripción

Especifica si las consultas UPDATE con varias coincidencias por fila devuelven un error.

## Ejemplo

```
SET error_on_nondeterministic_update TO true;

CREATE TABLE t1(x1 int, y1 int);

CREATE TABLE t2(x2 int, y2 int);

INSERT INTO t1 VALUES (1,10), (2,20), (3,30);

INSERT INTO t2 VALUES (2,40), (2,50);

UPDATE t1 SET y1=y2 FROM t2 WHERE x1=x2;

ERROR: Found multiple matches to update the same tuple.
```

# extra_float_digits

## Valores (predeterminados en negrita)

0, de -15 a 2

## Descripción

Establece la cantidad de dígitos que se muestran para los valores de punto flotante, incluidos los float4 y float8. El valor se agrega a la cantidad estándar de dígitos (FLT_DIG o DBL_DIG, según corresponda). El valor se puede establecer hasta 2 para incluir dígitos parcialmente significativos. Esto es especialmente útil para generar datos flotantes que deben restaurarse de forma exacta. También se puede establecer en negativo para suprimir dígitos no deseados.

## Ejemplo

En el siguiente ejemplo se establece `extra_float_digits` en `-2`. Primero, muestre la configuración de parámetros actual.

```
show all;
      name                      | setting
--------------------------+----------------
analyze_threshold_percent | 10
datestyle                 | ISO, MDY
extra_float_digits        | 2
query_group               | default
search_path               | $user, public
statement_timeout         | 0
timezone                  | UTC
wlm_query_slot_count      | 1
```

A continuación, defina el nuevo valor en `-2`.

```
set extra_float_digits to -2;
```

Por último, muestre la configuración de parámetros actualizada.

```
show all;
      name                      | setting
--------------------------+----------------
analyze_threshold_percent | 10
datestyle                 | ISO, MDY
extra_float_digits        | -2
query_group               | default
search_path               | $user, public
statement_timeout         | 0
```

```
timezone                  | UTC
wlm_query_slot_count      | 1
```

# interval_forbid_composite_literals

## Valores (predeterminados en negrita)

false, true

## Descripción

Una configuración de sesión que modifica el valor de un intervalo que contiene las partes YEAR TO MONTH y DAY TO SECOND.

Si `interval_forbid_composite_literals` es `true`, se devuelve un error si se encuentra un intervalo con las partes YEAR TO MONTH y DAY TO SECOND. Por ejemplo, el siguiente código SQL contiene INTERVAL DAY TO SECOND con las partes YEAR TO MONTH y DAY TO SECOND.

```
SELECT INTERVAL '1 year 1 day' DAY TO SECOND;
ERROR:  Interval Day To Second literal cannot contain year-month parts. Disable the GUC
 interval_forbid_composite_literals to suppress this error and silently discard the
 year-month part.
```

Si `interval_forbid_composite_literals` es `false`, Amazon Redshift suprime un error y trunca la parte YEAR TO MONTH de un valor INTERVAL DAY TO SECOND. Por ejemplo, el siguiente código SQL contiene INTERVAL DAY TO SECOND con las partes YEAR TO MONTH y DAY TO SECOND.

```
SET interval_forbid_composite_literals to "false";
SELECT INTERVAL '1 year 1 day' DAY TO SECOND;


 intervald2s
------------------------------
1 days 0 hours 0 mins 0.0 secs
```

# json_serialization_enable

## Valores (predeterminados en negrita)

false, true

## Descripción

Es la configuración de una sesión que modifica el comportamiento de serialización de JSON de los datos con formato ORC, JSON, Ion y Parquet. Si `json_serialization_enable` está configurado en `true`, todas las colecciones de nivel superior se serializan automáticamente a JSON y se devuelven como VARCHAR(65535). Las columnas no complejas no se ven afectadas ni se serializan. Dado que las columnas de colección se serializan como VARCHAR(65535), ya no se puede acceder directamente a sus subcampos anidados como parte de la sintaxis de la consulta (es decir, en la cláusula de filtrado). Si `json_serialization_enable` está configurado en `false`, las colecciones de nivel superior no se serializan a JSON. Para obtener más información acerca de la serialización de JSON anidada, consulte Serialización de datos JSON anidados y complejos.

# json_serialization_parse_nested_strings

## Valores (predeterminados en negrita)

false, true

## Descripción

Es la configuración de una sesión que modifica el comportamiento de serialización de JSON de los datos con formato ORC, JSON, Ion y Parquet. Cuando tanto `json_serialization_parse_nested_strings` como `json_serialization_enable` se configuran como true, los valores de cadena que se almacenan en tipos complejos (como mapas, estructuras o matrices) se analizan y escriben directamente en el resultado si son datos JSON válidos. Si `json_serialization_parse_nested_strings` está configurado como false, las cadenas dentro de tipos complejos anidados se serializan como cadenas JSON con escape. Para obtener más información, consulte Serialización de tipos complejos que contienen cadenas JSON.

# max_concurrency_scaling_clusters

## Valores (predeterminados en negrita)

1, de 0 a 10

## Descripción

Define el número máximo de clústeres de escalado de simultaneidad permitidos cuando el escalado de simultaneidad está habilitado. Aumente este valor si requiere un escalado de simultaneidad mayor. Utilice un valor menor para reducir el uso de clústeres de escalado de simultaneidad y los cargos facturados resultantes.

El número máximo de clústeres de escalado de simultaneidad es una cuota ajustable. Para obtener más información, consulte Cuotas de Amazon Redshift en la Guía de administración de Amazon Redshift.

# max_cursor_result_set_size

## Valores (predeterminados en negrita)

0 (de predeterminados a máximos) - 14 400 000 MB

## Descripción

El parámetro `max_cursor_result_set_size` ya no se usa. Para obtener más información acerca del tamaño de conjuntos de resultados de cursores, consulte Restricciones del cursor.

# mv_enable_aqmv_for_session

## Valores (predeterminados en negrita)

true, false

## Descripción

Especifica si Amazon Redshift puede realizar una reescritura automática de consultas de las vistas materializadas en el nivel de sesión.

# navegate_super_null_on_error

## Valores (predeterminados en negrita)

on (encendido), off (apagado)

## Descripción

Especifica que, cuando intenta recorrer un miembro inexistente de un objeto o elemento de una matriz, Amazon Redshift devuelve un valor NULL si la consulta se ejecuta en el modo relajado predeterminado.

# parse_super_null_on_error

## Valores (predeterminados en negrita)

off (apagado), on (encendido)

## Descripción

Especifica que, cuando Amazon Redshift intenta analizar un miembro inexistente de un objeto o elemento de una matriz, el servicio devuelve un valor NULL si la consulta se ejecuta en modo estricto.

# pg_federation_repeatable_read

## Valores (predeterminados en negrita)

true, false

## Descripción

Especifica el nivel de aislamiento de la transacción de consulta federada para los resultados de la base de datos PostgreSQL.

- Cuando `pg_federation_repeatable_read` es true, las transacciones federadas se procesan con la semántica de nivel de aislamiento de REPEATABLE READ. Esta es la opción predeterminada.

- Cuando pg_federation_repeatable_read es false, las transacciones federadas se procesan con la semántica de nivel de aislamiento de READ COMMITTED.

Para obtener más información, consulte los siguientes temas:

- Consideraciones al obtener acceso a los datos federados con Amazon Redshift.
- Administración de operaciones de escritura simultáneas.

## Ejemplos

El siguiente comando establece pg_federation_repeatable_read a on durante una sesión. El comando show muestra el valor del valor establecido.

```
set pg_federation_repeatable_read to on;
```

```
show pg_federation_repeatable_read;

pg_federation_repeatable_read
-----------------------------
on
```

# query_group

## Valores (predeterminados en negrita)

No hay valor predeterminado. El valor puede ser cualquier cadena de caracteres.

## Descripción

Aplica una etiqueta definida por el usuario a un grupo de consultas que se ejecutan durante la misma sesión. Esta etiqueta se captura en los registros de consultas. Puede utilizarla para limitar los resultados de las tablas STL_QUERY y STV_INFLIGHT, además de la vista de SVL_QLOG. Por ejemplo, puede aplicar una etiqueta independiente a cada consulta que desee ejecutar únicamente para identificar las consultas sin tener que buscar sus ID.

Este parámetro no existe en el archivo de configuración del servidor y debe establecerse durante el tiempo de ejecución con un comando SET. Aunque puede utilizar como etiqueta una cadena

de caracteres prolongada, la etiqueta se trunca en 30 caracteres en la columna LABEL de la tabla STL_QUERY y la vista SVL_QLOG (y en 15 caracteres en STV_INFLIGHT).

En el siguiente ejemplo, query_group se configura en **Monday** para luego ejecutar varias consultas con esa etiqueta.

```
set query_group to 'Monday';
SET
select * from category limit 1;
...
...
select query, pid, substring, elapsed, label
from svl_qlog where label ='Monday'
order by query;

query | pid  |        substring                   |  elapsed  | label
------+------+------------------------------------+-----------+--------
789   | 6084 | select * from category limit 1;    | 65468     | Monday
790   | 6084 | select query, trim(label) from ... | 1260327   | Monday
791   | 6084 | select * from svl_qlog where ..    | 2293547   | Monday
792   | 6084 | select count(*) from bigsales;     | 108235617 | Monday
...
```

# search_path

## Valores (predeterminados en negrita)

'$user', public, schema_names

Una lista de nombres de esquemas existentes separada por comas. Si '$user' está presente, se sustituye el esquema que tenga el mismo nombre que `SESSION_USER`; de lo contrario, se ignora.

## Descripción

Especifica el orden en el que se buscan los esquemas cuando se referencia un objeto (como una tabla o una función) con un nombre simple sin un componente de esquema:

- Las rutas de búsqueda no se admiten con esquemas externos ni tablas externas. Las tablas externas deben estar clasificadas explícitamente por un esquema externo. Los recursos compartidos de datos se comportan como datos externos y, por lo tanto, no son compatibles como rutas de búsqueda.

- Cuando se crean objetos sin un esquema de destino específico, estos se colocan en el primer esquema que aparece en la ruta de búsqueda. Si la ruta de búsqueda está vacía, el sistema devuelve un error.
- Cuando existen objetos con nombres idénticos en diferentes esquemas, se utiliza el que se encuentre primero en la ruta de búsqueda.
- Solo se podrá referenciar un objeto que no se encuentre en ninguno de los esquemas de la ruta de búsqueda mediante la especificación del esquema que lo contiene con un nombre (con puntos) calificado.
- El esquema del catálogo del sistema, pg_catalog, se busca siempre. Si se menciona en la ruta, se busca en el orden especificado. Si no, se busca antes que a cualquiera de los elementos de la ruta.
- El esquema de tablas temporales de la sesión actual, pg_temp_nnn, se busca siempre, si existe. Puede aparecer explícitamente en la ruta con el alias pg_temp. Si no aparece en la ruta, se busca primero (incluso antes que pg_catalog). No obstante, el esquema temporal solo se busca por nombres de relación (tablas, vistas). No se busca por nombres de funciones.

## Ejemplo

En el siguiente ejemplo, se crea el esquema ENTERPRISE y se establece el search_path para el esquema nuevo.

```
create schema enterprise;
set search_path to enterprise;
show search_path;

 search_path
-------------
 enterprise
(1 row)
```

En el siguiente ejemplo, se agrega el esquema ENTERPRISE al search_path predeterminado.

```
set search_path to '$user', public, enterprise;
show search_path;

         search_path
-----------------------------
 "$user", public, enterprise
```

```
(1 row)
```

En el siguiente ejemplo, se agrega la tabla FRONTIER al esquema ENTERPRISE.

```
create table enterprise.frontier (c1 int);
```

Cuando la tabla PUBLIC.FRONTIER se crea en la misma base de datos, y el usuario no especifica el nombre del esquema en una consulta, PUBLIC.FRONTIER tiene prioridad sobre ENTERPRISE.FRONTIER:

```
create table public.frontier(c1 int);
insert into enterprise.frontier values(1);
select * from frontier;

frontier
----
(0 rows)

select * from enterprise.frontier;

c1
----
1
(1 row)
```

# spectrum_enable_pseudo_columns

## Valores (predeterminados en negrita)

true, false

## Descripción

Puede desactivar la creación de pseudocolumnas para una sesión estableciendo el parámetro de configuración `spectrum_enable_pseudo_columns` en `false`.

## Ejemplo

El siguiente comando desactiva la creación de pseudocolumnas para una sesión.

```
set spectrum_enable_pseudo_columns to false;
```

# enable_spectrum_oid

## Valores (predeterminados en negrita)

true, false

## Descripción

Puede desactivar solo la pseudocolumna `$spectrum_oid` configurando el parámetro de configuración `enable_spectrum_oid` en `false`.

## Ejemplo

El siguiente comando desactiva la pseudocolumna `$spectrum_oid` configurando el parámetro de configuración `enable_spectrum_oid` en `false`.

```
set enable_spectrum_oid to false;
```

# spectrum_query_maxerror

## Valores (predeterminados en negrita)

-1, entero

## Descripción

Aprenda a especificar un número entero para indicar el número máximo de errores que se pueden aceptar antes de cancelar la consulta. Un valor negativo desactiva el control máximo de datos de errores. Los resultados se han registrado en SVL_SPECTRUM_SCAN_ERROR.

## Ejemplo

En el siguiente ejemplo, se da por hecho que hay datos de ORC que contienen caracteres sobrantes y no válidos. La definición de columna de `my_string` especifica una longitud de 3 caracteres. A continuación se indican los datos de muestra de este ejemplo.

```
my_string
---------------
abcdef
gh
ab
```

Los siguientes comandos establecen el número máximo de errores en 1 y realizan la consulta.

```
set spectrum_query_maxerror to 1;
SELECT my_string FROM orc_data;
```

La consulta se detiene y los resultados se registran en SVL_SPECTRUM_SCAN_ERROR.

# statement_timeout

## Valores (predeterminados en negrita)

0 (desactiva la limitación), x milisegundos

## Descripción

Detiene cualquier instrucción que tome el control sobre la cantidad de milisegundos especificada.

El valor `statement_timeout` es la cantidad máxima de tiempo que puede ejecutarse una consulta antes de que Amazon Redshift la termine. Este tiempo incluye el tiempo de planificación, el tiempo que se espera en cola en la administración de cargas de trabajo (WLM) y el tiempo de ejecución. Compare este tiempo con el tiempo de espera de WLM (max_execution_time) y una QMR (query_execution_time), que solo incluyen el tiempo de ejecución.

Si también se especifica el tiempo de espera de WLM (max_execution_time) como parte de una configuración de WLM, se utilizan los valores de statement_timeout y max_execution_time inferiores. Para obtener más información, consulte Tiempo de espera de WLM.

## Ejemplo

Dado que la siguiente consulta tarda más de 1 milisegundo, se agota el tiempo de espera y la consulta se cancela.

```
set statement_timeout = 1;
```

```
select * from listing where listid>5000;
ERROR:  Query (150) canceled on user's request
```

# stored_proc_log_min_messages

## Valores (predeterminados en negrita)

LOG, INFO, NOTICE, WARNING, EXCEPTION

## Descripción

Especifica el nivel mínimo de registro de los mensajes de procedimiento almacenados generados. Se registran mensajes en el nivel especificado o por encima de este. El valor predeterminado es LOG (todos los mensajes se registran). El orden de los niveles de registro de más alto a más bajo es el siguiente:

1. EXCEPTION
2. ADVERTENCIA
3. NOTICE
4. INFO
5. LOG

Por ejemplo, si especifica un valor de NOTICE, los mensajes solo se registran para NOTICE, WARNING y EXCEPTION.

# timezone

## Valores (predeterminados en negrita)

UTC, zona horaria

## Sintaxis

```
SET timezone { TO | = } [ time_zone | DEFAULT ]
```

```
SET time zone [ time_zone | DEFAULT ]
```

## Descripción

Establece la zona horaria para la sesión actual. La zona horaria puede ser el desplazamiento de la hora universal coordinada (UTC) o el nombre de una zona horaria.

> **Note**
>
> No puede establecer el parámetro de configuración `timezone` utilizando un grupo de parámetros de clúster. La zona horaria se puede establecer solo para la sesión actual con un comando SET. Para establecer la zona horaria para todas las sesiones ejecutadas por un usuario específico de la base de datos, utilice el comando ALTER USER. ALTER USER … SET TIMEZONE cambia la zona horaria para las sesiones posteriores, no para la actual.

Cuando establece la zona horaria mediante el comando `SET timezone` (una palabra) con `TO` o `=`, puede especificar time_zone como el nombre de una zona horaria, un desplazamiento con formato estilo POSIX o un desplazamiento con formato ISO-8601, tal y como se muestra a continuación.

```
SET timezone { TO | = } time_zone
```

Si establece la zona horaria mediante el comando SET time zone sin `TO` ni `=`, puede especificar time_zone utilizando INTERVAL y también el nombre de una zona horaria, un desplazamiento con formato de estilo POSIX o un desplazamiento con formato ISO-8601, tal y como se muestra a continuación.

```
SET time zone time_zone
```

## Formatos de zonas horarias

Amazon Redshift admite los siguientes formatos de zonas horarias:

- Nombre de zona horaria
- INTERVAL
- Especificación de zona horaria estilo POSIX

- Desplazamiento ISO-8601

Dado que las abreviaturas de zonas horarias, como PST o PDT, se definen como un desplazamiento fijo de UTC y no incluyen reglas de horario de verano, el comando SET no admite las abreviaturas de zonas horarias.

Para obtener más detalles acerca de los formatos de zonas horarias, consulte lo siguiente.

Nombre de zona horaria: nombre completo de la zona horaria; por ejemplo, America/New_York. Los nombres completos de zonas horarias pueden incluir reglas de horario de verano.

A continuación, se muestran ejemplos de nombres de zonas horarias:

- Etc/Greenwich
- America/New_York
- CST6CDT
- GB

> **Note**
>
> Varios nombres de zonas horarias, como EST, MST, NZ y UCT, también son abreviaturas.

Para ver una lista de nombres de zonas horarias válidos, ejecute el siguiente comando.

```
select pg_timezone_names();
```

INTERVAL: desplazamiento de UTC. Por ejemplo, PST es –8:00 o –8 horas.

A continuación se muestran ejemplos de desplazamientos de zonas horarias de INTERVAL:

- –8:00
- –8 horas
- 30 minutos

Formato de estilo POSIX: especificación de zona horaria en la forma STDoffset o STDoffsetDST, donde STD es la abreviatura de la zona horaria, offset es el desplazamiento numérico en horas

hacia el oeste desde la UTC y DST es una abreviatura opcional del horario de verano de la zona. Se supone que el horario de verano adelanta una hora, en función del desplazamiento dado.

Los formatos de zona horaria estilo POSIX usan desfases positivos al oeste de Greenwich, al contrario que la convención de la norma ISO 8601, que usa desfases positivos al este de Greenwich.

A continuación, se muestran algunos ejemplos de zonas horarias estilo POSIX:

- PST8
- PST8PDT
- EST5
- EST5EDT

> **Note**
>
> Amazon Redshift no valida las especificaciones de zona horaria con estilo POSIX, por lo que es posible establecer la zona horaria en un valor no válido. Por ejemplo, el siguiente comando no devuelve un mensaje de error, a pesar de que establece la zona horaria en un valor no válido.
>
> ```
> set timezone to ‘xxx36’;
> ```

Desplazamiento ISO-8601: desplazamiento desde la UTC con el formato `±[hh]:[mm]`.

A continuación, se muestran ejemplos de desplazamientos ISO-8601:

- -8:00
- +7:30

## Ejemplos

En el siguiente ejemplo, se establece la zona horaria para la sesión actual en Nueva York.

```
set timezone = 'America/New_York';
```

En el siguiente ejemplo, se establece la zona horaria para la sesión actual en UTC–8 (PST).

```
set timezone to '-8:00';
```

En el siguiente ejemplo, se utiliza INTERVAL para establecer la zona horaria en PST.

```
set timezone interval '-8 hours'
```

En el siguiente ejemplo, se restablece la zona horaria para la sesión actual en la zona horaria predeterminada del sistema (UTC).

```
set timezone to default;
```

Para establecer la zona horaria para el usuario de la base de datos, utilice una instrucción ALTER USER … SET. En el siguiente ejemplo, se establece la zona horaria para el usuario de la base de datos en Nueva York. El nuevo valor se mantiene para el usuario para todas las sesiones posteriores.

```
ALTER USER dbuser SET timezone to 'America/New_York';
```

# use_fips_ssl

## Valores (predeterminados en negrita)

true, false

## Descripción

Un valor de grupo de parámetros que especifica si se utiliza el modo SSL compatible con FIPS. Si `use_fips_ssl` es `true`, se usa el modo SSL compatible con FIPS. Si `use_fips_ssl` es `false`, no se usa el modo SSL compatible con FIPS. Para obtener más información, consulte Configuración de las opciones de seguridad para las conexiones en la Guía de administración de Amazon Redshift.

Para configurar los parámetros de un clúster aprovisionado de Amazon Redshift, consulte Acerca de los grupos de parámetros en la Guía de administración de Amazon Redshift. Para configurar los parámetros de Redshift sin servidor, consulte Configuración de una conexión SSL compatible con FIPS en Amazon Redshift sin servidor en la Guía de administración de Amazon Redshift y CreateWorkgroup o UpdateWorkgroup en la Referencia de la API de Redshift sin servidor.

# wlm_query_slot_count

## Valores (predeterminados en negrita)

1, de 1 a 50; no debe exceder el número de slots disponibles (nivel de concurrencia) de la clase de servicio.

## Descripción

Establece la cantidad de slots de consultas que utiliza una consulta.

La administración de cargas de trabajo (WLM) reserva slots en una clase de servicio de acuerdo con el nivel de simultaneidad definido para la cola. Por ejemplo, si el nivel de simultaneidad se establece en 5, la clase de servicio tendrá 5 slots. WLM asigna la memoria disponible para una clase de servicio a cada slot equitativamente. Para obtener más información, consulte Administración de la carga de trabajo.

> **Note**
>
> Si el valor de wlm_query_slot_count es mayor que la cantidad de slots disponibles (nivel de simultaneidad) para la clase de servicio, la consulta produce un error. Si detecta un error, disminuya wlm_query_slot_count a un valor permitido.

Para las operaciones en las que el rendimiento se ve afectado en gran medida por la cantidad de memoria asignada, como la limpieza, el aumento del valor de wlm_query_slot_count puede mejorar el rendimiento. En concreto, para los comandos de limpieza, examine el registro correspondiente en la vista SVV_VACUUM_SUMMARY. Si observa valores altos (cercanos o mayores a 100) para sort_partitions y merge_increments en la vista SVV_VACUUM_SUMMARY, considere aumentar el valor para wlm_query_slot_count la próxima vez que ejecute Vacuum en esa tabla.

El aumento del valor de wlm_query_slot_count limita la cantidad de consultas simultáneas que se pueden ejecutar. Por ejemplo, suponga que la clase de servicio tiene un nivel de simultaneidad de 5 y wlm_query_slot_count está establecido en 3. Mientras una consulta se ejecuta dentro de la sesión con wlm_query_slot_count establecido en 3, se puede ejecutar un máximo de 2 consultas simultáneas más dentro de la misma clase de servicio. Las consultas posteriores esperan en la cola hasta que se completen las consultas actualmente en ejecución y se liberen los slots.

# Ejemplos

Utilice el comando SET para establecer el valor de wlm_query_slot_count para el tiempo que dure la sesión actual.

```
set wlm_query_slot_count to 3;
```

# Historial del documento

> **Note**
>
> Para obtener una descripción de las nuevas características de Amazon Redshift, consulte Novedades.

En la siguiente tabla, se describen los cambios importantes de la documentación de la Guía para desarrolladores de bases de datos de Amazon Redshift después de mayo de 2018. Para recibir notificaciones sobre los cambios en esta documentación, puede suscribirse a una fuente RSS.

Versión de API: 2012-12-01

Para ver una lista con los cambios realizados en la Guía de administración de Amazon Redshift, consulte Historial de revisión de la Guía de administración de Amazon Redshift.

Para obtener más información acerca de las nuevas características, incluida una lista de correcciones y los números de versión del clúster asociado, visite el historial de versiones del clúster.

| Cambio | Descripción | Fecha |
| --- | --- | --- |
| Compatibilidad con geometrías espaciales en 3D y 4D y nuevas funciones espaciales | Ahora puede utilizar funciones espaciales adicionales. Por otro lado, se agrega compatibilidad con geometría en 3D y 4D a algunas funciones. | 19 de agosto de 2021 |
| Compatibilidad con la codificación de compresión de columnas para la optimización automática de tablas | Puede especificar la opción ENCODE AUTO para una tabla con el objetivo de administrar automáticamente la codificación de compresión para todas las columnas de la tabla. | 3 de agosto de 2021 |
| Compatibilidad con varias instrucciones SQL o | Ahora puede ejecutar varias instrucciones SQL o una | 28 de julio de 2021 |

| | | |
|---|---|---|
| una instrucción SQL con parámetros por medio de la API de datos de Amazon Redshift | instrucción con parámetros con la API de datos de Amazon Redshift. | |
| Compatibilidad con la intercalación que no distingue entre mayúsculas y minúsculas con anulaciones de nivel de columna | Ahora puede utilizar la cláusula COLLATE dentro de una instrucción CREATE DATABASE para especificar la intercalación predeterminada. | 24 de junio de 2021 |
| Compatibilidad con el uso compartido de datos entre cuentas | Ahora puede compartir datos entre Cuentas de AWS. | 30 de abril de 2021 |
| Compatibilidad con consultas jerárquicas de datos con CTE recursivas | Ahora puede utilizar una expresión de tabla común (CTE) recursiva en SQL. | 29 de abril de 2021 |
| Compatibilidad con consultas entre bases de datos | Ahora puede consultar datos en las bases de datos de un clúster. | 10 de marzo de 2021 |
| Compatibilidad con el control de acceso pormenorizado en los comandos COPY y UNLOAD | Ahora puede otorgar el privilegio de ejecutar los comandos COPY y UNLOAD a usuarios y grupos específicos del clúster de Amazon Redshift para crear una política de control de acceso más detallada. | 12 de enero de 2021 |
| Compatibilidad con los datos JSON nativos y semiestructurados | Ahora, puede definir el tipo de datos SUPER. | 9 de diciembre de 2020 |

| | | |
|---|---|---|
| Compatibilidad con las consultas federadas a MySQL | Ahora, puede escribir una consulta federada en un motor MySQL compatible. | 9 de diciembre de 2020 |
| Compatibilidad con el uso compartido de datos | Ahora, puede compartir datos entre clústeres de Amazon Redshift. | 9 de diciembre de 2020 |
| Compatibilidad con la optimización automática de tablas | Ahora, puede definir claves de ordenación y distribución automáticas. | 9 de diciembre de 2020 |
| Compatibilidad con Amazon Redshift ML | Ahora, puede crear, formar e implementar modelos de machine learning (ML). | 8 de diciembre de 2020 |
| Compatibilidad con la actualización automática y la reescritura de consultas de las vistas materializadas | Ahora, puede mantener actualizadas las vistas materializadas con la actualización automática, y el rendimiento de las consultas se puede mejorar con la reescritura automática. | 11 de noviembre de 2020 |
| Compatibilidad con los tipos de datos TIME y TIMETZ | Ahora, puede crear tablas con tipos de datos TIME y TIMETZ. El tipo de datos TIME almacena la hora del día sin información de la zona horaria, y TIMETZ almacena la hora del día con información de la zona horaria. | 11 de noviembre de 2020 |
| Compatibilidad con la tokenización y las UDF de Lambda | Ahora, puede escribir las UDF de Lambda para habilitar la tokenización externa de datos. | 26 de octubre de 2020 |

| | | |
|---|---|---|
| Compatibilidad con la modificación de la codificación de columnas de una tabla | Ahora, puede modificar la codificación de columnas de una tabla. | 20 de octubre de 2020 |
| Compatibilidad con la realización de consultas entre bases de datos | Amazon Redshift ahora puede consultar distintas bases de datos de un clúster. | 15 de octubre de 2020 |
| Compatibilidad con los bocetos de HyperLogLog | Amazon Redshift ahora puede almacenar y procesar bocetos de HyperLogLog. | 2 de octubre de 2020 |
| Compatibilidad con Apache Hudi y Delta Lake | Se trata de mejoras en la creación de tablas externas para Redshift Spectrum. | 24 de septiembre de 2020 |
| Compatibilidad con las mejoras en la consulta de datos espaciales | Las mejoras incluyen la carga de un shapefile y varias funciones SQL espaciales nuevas. | 15 de septiembre de 2020 |
| Compatibilidad de las vistas materializadas con tablas externas | Puede crear vistas materializadas en Amazon Redshift que referencien orígenes de datos externos. | 19 de junio de 2020 |
| Compatibilidad con la escritura en tablas externas | Puede escribir en tablas externas ejecutando CREATE EXTERNAL TABLE AS SELECT para escribir en una nueva tabla externa o INSERT INTO para insertar datos en una tabla externa existente. | 8 de junio de 2020 |
| Compatibilidad con los controles de almacenamiento de esquemas | Se han actualizado comandos y vistas que administran controles de almacenamiento para los esquemas. | 2 de junio de 2020 |

| | | |
|---|---|---|
| Compatibilidad con la disponibilidad general de consultas federadas | Se ha actualizado la información acerca de cómo consultar datos con consultas federadas . | 16 de abril de 2020 |
| Compatibilidad con nuevas funciones espaciales | Se han agregado descripciones sobre nuevas funciones espaciales. | 2 de abril de 2020 |
| Compatibilidad con la disponibilidad general de las vistas materializadas | Las vistas materializadas generalmente están disponibles a partir de la versión de clúster 1.0.13059. | 19 de febrero de 2020 |
| Compatibilidad con privilegios de nivel de columna | Los privilegios de nivel de columna están disponibles a partir de la versión de clúster 1.0.13059. | 19 de febrero de 2020 |
| ALTER TABLE | Puede utilizar un comando ALTER TABLE con la cláusula ALTER DISTSTYLE ALL para cambiar el estilo de distribución de una tabla. | 11 de febrero de 2020 |
| Compatibilidad con consultas federadas | Se ha actualizado la guía para describir la consulta federada con un comando CREATE EXTERNAL SCHEMA actualizado. | 3 de diciembre de 2019 |
| Compatibilidad con la exportación de lago de datos | Se ha actualizado la guía para describir los nuevos parámetros del comando UNLOAD. | 3 de diciembre de 2019 |
| Compatibilidad con los datos espaciales | Se ha actualizado la guía para describir la compatibilidad con los datos espaciales. | 21 de noviembre de 2019 |

| | | |
|---|---|---|
| Soporte para la nueva consola | Actualizamos la guía para describir la nueva consola de Amazon Redshift. | 11 de noviembre de 2019 |
| Compatibilidad con la clasificación automática de tablas | Amazon Redshift puede ordenar de manera automática los datos de las tablas. | 7 de noviembre de 2019 |
| Compatibilidad con la opción VACUUM BOOST | Puede usar la opción BOOST al realizar una limpieza de tablas. | 7 de noviembre de 2019 |
| Compatibilidad con columnas IDENTITY predeterminadas | Puede crear tablas con columnas IDENTITY predeterminadas. | 19 de septiembre de 2019 |
| Compatibilidad con la codificación de compresión AZ64 | Puede codificar algunas columnas con la codificación de compresión AZ64. | 19 de septiembre de 2019 |
| Compatibilidad para prioridad de consultas | Puede establecer la prioridad de consultas de una cola de WLM automática. | 22 de agosto de 2019 |
| Compatibilidad con AWS Lake Formation | Puede utilizar un catálogo de datos de Lake Formation con Amazon Redshift Spectrum. | 8 de agosto de 2019 |
| COMPUPDATE PRESET | Puede utilizar un comando COPY con COMPUPDATE PRESET para permitir que Amazon Redshift elija la codificación de compresión. | 13 de junio de 2019 |

| | | |
|---|---|---|
| ALTER COLUMN | Puede utilizar un comando ALTER TABLE con ALTER COLUMN para aumentar el tamaño de una columna VARCHAR. | 22 de mayo de 2019 |
| Compatibilidad con procedimientos almacenados | Puede definir procedimientos almacenados de PL/pgSQL en Amazon Redshift. | 24 de abril de 2019 |
| Compatibilidad para una configuración de administración de carga de trabajo automática (WLM) | Puede permitir que Amazon Redshift se ejecute con la WLM automática. | 24 de abril de 2019 |
| UNLOAD en Zstandard | Puede utilizar el comando UNLOAD para aplicar la compresión Zstandard a archivos de texto y de valores separados por comas (CSV) descargados en Amazon S3. | 3 de abril de 2019 |
| Escalado de simultaneidad | Cuando el escalado de simultaneidad está habilitado, Amazon Redshift agrega capacidad del clúster de manera automática si se necesita para procesar un aumento de las consultas de lectura simultáneas. | 21 de marzo de 2019 |
| UNLOAD en CSV | Puede utilizar el comando UNLOAD para descargar en un archivo con formato de texto CSV. | 13 de marzo de 2019 |

| | | |
|---|---|---|
| Estilo de distribución AUTO | Para habilitar la distribución automática, puede especificar el estilo de distribución AUTO con una instrucción CREATE TABLE. Cuando habilita la distribución automática, Amazon Redshift asigna un estilo de distribución óptimo basado en los datos de la tabla. El cambio en la distribución se produce en segundo plano y tarda pocos segundos. | 23 de enero de 2019 |
| COPY de Parquet admite SMALLINT | COPY ahora admite la carga desde archivos con formato Parquet en columnas que utilizan el tipo de datos SMALLINT. Para obtener más información, consulte Uso de COPY con formatos de datos de columnas. | 2 de enero de 2019 |
| DROP EXTERNAL DATABASE | Puede eliminar una base de datos externa incluyendo la cláusula DROP EXTERNAL DATABASE con un comando DROP SCHEMA. | 3 de diciembre de 2018 |
| Operación UNLOAD entre regiones | Puede realizar una operación UNLOAD en un bucket de Amazon S3 en otra región de AWS especificando el parámetro REGION. | 31 de octubre de 2018 |

| | | |
|---|---|---|
| Eliminación de limpieza automática | Amazon Redshift ejecuta de manera automática una operación VACUUM DELETE en segundo plano, por lo que rara vez tendrá que ejecutar una limpieza DELETE ONLY. Amazon Redshift programa la operación VACUUM DELETE para que se ejecute durante periodos de carga reducida y pone en pausa la operación durante periodos de carga elevada. | 31 de octubre de 2018 |
| Distribución automática | Si no especifica un estilo de distribución con una instrucción CREATE TABLE, Amazon Redshift asigna un estilo de distribución óptimo basado en los datos de la tabla. El cambio en la distribución se produce en segundo plano y tarda pocos segundos. | 31 de octubre de 2018 |
| Control del acceso detallado para AWS Glue Data Catalog | Ahora, puede especificar niveles de acceso a los datos almacenados en el AWS Glue Data Catalog. | 15 de octubre de 2018 |
| UNLOAD con tipos de datos | Puede especificar la opción MANIFEST VERBOSE con un comando UNLOAD para agregar metadatos al archivo de manifiesto, incluyendo los nombres y tipos de datos de columnas, tamaños de archivo y recuentos de filas. | 10 de octubre de 2018 |

| | | |
|---|---|---|
| Agregar múltiples particiones en una única instrucción ALTER TABLE | Para tablas externas de Redshift Spectrum, puede combinar múltiples cláusulas PARTITION en una única instrucción ALTER TABLE ADD. Para obtener más información, consulte los Ejemplos de modificación de tabla externa. | 10 de octubre de 2018 |
| UNLOAD con encabezado | Puede especificar la opción HEADER con un comando UNLOAD para añadir una línea de encabezado que contenga los nombres de columna al comienzo de cada archivo de salida. | 19 de septiembre de 2018 |
| Nuevas tablas y vistas del sistema | Se ha añadido documentación de SVL_S3Retries, SVL_USER_INFO y STL_DISK_FULL_DIAG. | 31 de agosto de 2018 |
| Compatibilidad con datos anidados en Amazon Redshift Spectrum | Ahora puede consultar datos anidados almacenados en tablas de Amazon Redshift Spectrum. Para obtener más información, consulte el Tutorial: Consultar datos anidados con Amazon Redshift Spectrum. | 8 de agosto de 2018 |

| | | |
|---|---|---|
| SQA "on" de manera predeterminada | La aceleración de consultas cortas (SQA) está ahora habilitada de manera predeterminada para todos los nuevos clústeres. La aceleración de consultas cortas utiliza el aprendizaje automático para ofrecer un mayor desempeño, resultados más rápidos y una mejor previsibilidad de los tiempos de ejecución de las consultas. Para obtener más información, consulte Aceleración de consultas cortas. | 8 de agosto de 2018 |
| Amazon Redshift Advisor | Ahora, con Amazon Redshift Advisor, puede obtener recomendaciones adaptadas a su caso sobre cómo mejorar el rendimiento del clúster y reducir los costos operativos. Para obtener más información consulte Amazon Redshift Advisor. | 26 de julio de 2018 |
| Referencia de alias inmediata | Ahora puede hacer referencia a una expresión con alias inmediatamente después de definirla. Para obtener más información consulte la sección Lista SELECT. | 18 de julio de 2018 |

| | | |
|---|---|---|
| Especificar el tipo de compresión al crear una tabla externa | Ahora puede especificar el tipo de compresión al crear una tabla externa con Amazon Redshift Spectrum. Para obtener más información, consulte Creación de tablas externas. | 27 de junio de 2018 |
| PG_LAST_UNLOAD_ID | Se ha añadido documentación para una nueva función de información del sistema: PG_LAST_UNLOAD_ID. Para obtener más información, consulte PG_LAST_UNLOAD_ID. | 27 de junio de 2018 |
| ALTER TABLE RENAME COLUMN | ALTER TABLE ahora permite cambiar el nombre de las columnas de las tablas externas. Para obtener más información, consulte los Ejemplos de modificación de tabla externa. | 7 de junio de 2018 |

## Actualizaciones anteriores

En la siguiente tabla, se describen los cambios importantes que se han realizado en cada una de las versiones de la Guía para desarrolladores de bases de datos de Amazon Redshift antes de junio de 2018.

| Cambio | Descripción | Fecha de modificación |
|---|---|---|
| COPY de Parquet incluye SMALLINT | COPY ahora admite la carga desde archivos con formato Parquet en columnas que utilizan el tipo de datos SMALLINT. Para obtener más información, | 2 de enero de 2019 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| | consultar Uso de COPY con formatos de datos de columnas | |
| Uso de COPY con formatos de columnas | Ahora, COPY permite cargar datos desde archivos de Amazon S3 que usan los formatos de columnas Parquet y ORC. Para obtener más información, consultar Uso de COPY con formatos de datos de columnas | 17 de mayo de 2018 |
| Tiempo de ejecución máximo dinámico de SQA | De manera predeterminada, ahora la administración de cargas de trabajo (WLM) asigna dinámicamente un valor para el tiempo de ejecución máximo de aceleración de consultas cortas (SQA) en función del análisis de la carga de trabajo del clúster. Para obtener más información, consulte Tiempo máximo de ejecución de consultas cortas. | 17 de mayo de 2018 |
| Nueva columna en STL_LOAD_COMMITS | La tabla del sistema STL_LOAD_COMMITS tiene una nueva columna: `file_format` . | 10 de mayo de 2018 |
| Nuevas columnas en STL_HASHJOIN y otras tablas de registro del sistema | La tabla del sistema STL_HASHJOIN tiene tres nuevas columnas: `hash_segment` , `hash_step` y `checksum`. Además, hemos incorporado una `checksum` a STL_MERGEJOIN, STL_NESTLOOP, STL_HASH, STL_SCAN, STL_SORT, STL_LIMIT y STL_PROJECT. | 17 de mayo de 2018 |
| Nuevas columnas en STL_AGGR | La tabla del sistema STL_AGGR tiene dos nuevas columnas: `resizes` y `flushable` . | 19 de abril de 2018 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Opciones nuevas para las funciones REGEX | Para las funciones REGEXP_INSTR y REGEXP_SUBSTR, ahora puede especificar cuál de los valores coincidentes se debe utilizar y si desea que la función distinga entre mayúsculas y minúsculas. REGEXP_INSTR también le permite especificar si se debe a devolver la posición del primer carácter de la coincidencia o la posición del primer carácter situado a continuación del final de la coincidencia. | 22 de marzo de 2018 |
| Columnas nuevas en las tablas del sistema | Se han añadido las columnas tombstonedblocks, tossedblocks y batched_by a la tabla del sistema STL_COMMIT_STATS. Se ha añadido la columna localslice a la vista del sistema STV_SLICES. | 22 de marzo de 2018 |
| Adición y eliminación de columnas en las tablas externas | Ahora, ALTER TABLE admite ADD COLUMN y DROP COLUMN para las tablas externas de Amazon Redshift Spectrum. | 22 de marzo de 2018 |
| Regiones nuevas de AWS de Redshift Spectrum | Redshift Spectrum ya está disponible en las regiones de Mumbai and São Paulo. Para obtener una lista de las regiones admitidas, consulte Regiones de Amazon Redshift Spectrum. | 22 de marzo de 2018 |
| Se ha aumentado el límite de tablas a 20 000 | El número máximo de tablas ahora es de 20 000 para los tipos de nodos de clúster 8xlarge. El límite para los tipos de nodos large y xlarge es de 9 900. Para obtener más información, consulte Límites y cuotas. | 13 de marzo de 2018 |
| Compatibilidad de Redshift Spectrum con JSON e Ion | Redshift Spectrum permite hacer referencia a archivos con datos escalares en los formatos de datos JSON o Ion. Para obtener más información, consulte CREATE EXTERNAL TABLE. | 26 de febrero de 2018 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Encadenamiento de roles de IAM en Redshift Spectrum | Puede encadenar roles de AWS Identity and Access Management (IAM) de manera que el clúster pueda asumir otros roles que no estén adjuntos al clúster, incluidos los roles que pertenezcan a otra cuenta de AWS. Para obtener más información, consulte Encadenamiento de roles de IAM en Amazon Redshift Spectrum. | 1 de febrero de 2018 |
| ADD PARTITION admite IF NOT EXISTS | La cláusula ADD PARTITION de ALTER TABLE ahora admite una opción IF NOT EXISTS. Para obtener más información, consulte ALTER TABLE. | 11 de enero de 2018 |
| Tipo de datos DATE para las tablas externas | Ahora, las tablas externas de Redshift Spectrum admiten el tipo de datos DATE. Para obtener más información, consulte CREATE EXTERNAL TABLE. | 11 de enero de 2018 |
| Regiones nuevas de AWS de Redshift Spectrum | Redshift Spectrum ya está disponible en las regiones de Singapur, Sídney, Seúl y Fráncfort. Para obtener una lista de las regiones de AWS admitidas, consulte Regiones de Amazon Redshift Spectrum. | 16 de noviembre de 2017 |
| Aceleración de consultas cortas en la administración de cargas de trabajo (WLM) de Amazon Redshift | La aceleración de consultas cortas (SQA) da prioridad a una serie de consultas seleccionadas que se ejecutan rápidamente frente a consultas que tardan más en ejecutarse. SQA ejecuta las consultas cortas en un espacio dedicado, de forma que estas consultas no tienen que esperar en las colas detrás de otras consultas más largas. Con SQA, las consultas cortas se ejecutan con mayor rapidez y los usuarios tardan menos en ver los resultados. Para obtener más información, consulte Aceleración de consultas cortas. | 16 de noviembre de 2017 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| WLM reasigna las consultas transferidas | En lugar de cancelar y reiniciar las consultas transferidas, la administración de cargas de trabajo (WLM) de Amazon Redshift ahora reasigna las consultas que reúnan los requisitos necesarios a una nueva cola. Cuando WLM reasigna una consulta, la mueve a una nueva cola y continúa la ejecución, lo que ahorra tiempo y recursos del sistema. Las consultas transferidas que no cumplen los requisitos necesarios para reasignarlas se reinician o se cancelan. Para obtener más información, consulte Salto de cola de consultas de WLM. | 16 de noviembre de 2017 |
| Acceso de los usuarios a los registros del sistema | En la mayoría de las tablas que están visibles para los usuarios, de manera predeterminada, los usuarios normales no pueden ver las filas que generan otros usuarios. Para permitir que un usuario normal pueda ver todas las filas incluidas en tablas visibles para los usuarios, incluidas las filas generadas por otro usuario, ejecute ALTER USER o CREAR USUARIO y establezca el parámetro SYSLOG ACCESS en UNRESTRICTED. | 16 de noviembre de 2017 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Almacenamiento en caché de los resultados | Con Almacenamiento en caché de los resultados, cuando se ejecuta una consulta, Amazon Redshift guarda el resultado en el almacenamiento de caché. Si esa consulta vuelve a ejecutarse, Amazon Redshift revisa el almacenamiento de caché en busca de una copia válida de los resultados de dicha consulta. Si se encuentra alguna coincidencia en la memoria caché de resultados, Amazon Redshift utiliza este resultado y no ejecuta la consulta. El almacenamiento en la caché de resultados está activado de manera predeterminada. Para desactivarlo, establezca el parámetro de configuración enable_result_cache_for_session en `off`. | 16 de noviembre de 2017 |
| Funciones de metadatos de columna | PG_GET_COLS y PG_GET_LATE_BINDING_VIEW_COLS devuelven los metadatos de columna que corresponden a las tablas, las vistas y las vistas de enlace de tiempo de ejecución de Amazon Redshift. | 16 de noviembre de 2017 |
| Salto de cola de WLM para CTAS | Ahora, la administración de cargas de trabajo (WLM) de Amazon Redshift admite el salto de cola de las consultas para las instrucciones CREATE TABLE AS (CTAS), así como las consultas de solo lectura, como las instrucciones SELECT. Para obtener más información, consulte Salto de cola de consultas de WLM. | 19 de octubre de 2017 |
| Archivos de manifiesto de Amazon Redshift Sp | Cuando crea una tabla externa de Redshift Spectrum, puede especificar un archivo de manifiesto que enumere las ubicaciones de los archivos de datos en Amazon S3. Para obtener más información, consulte CREATE EXTERNAL TABLE. | 19 de octubre de 2017 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Regiones nuevas de AWS de Amazon Redshift Sp | Redshift Spectrum ya está disponible en las regiones de UE (Irlanda) y Asia Pacífico (Tokio). Para obtener una lista de las regiones de AWS admitidas, consulte Limitaciones de Amazon Redshift Spectrum. | 19 de octubre de 2017 |
| Formatos de archivos agregados a Amazon Redshift Sp | Ahora puede crear tablas externas de Redshift Spectrum basadas en los formatos de archivo de datos Regex, OpenCSV y Avro. Para obtener más información, consulte CREATE EXTERNAL TABLE. | 5 de octubre de 2017 |
| Pseudocolumnas para tablas externas de Amazon Redshift Sp | Puede seleccionar las pseudocolumnas `$path` y `$size` en una tabla externa de Redshift Spectrum para ver la ubicación y el tamaño de los archivos de datos que se referencian en Amazon S3. Para obtener más información, consulte Pseudocolumnas. | 5 de octubre de 2017 |
| Funciones para validar JSON | Puede usar las funciones IS_VALID_JSON y IS_VALID_JSON_ARRAY para comprobar si hay un formato JSON válido. Ahora, las otras funciones JSON tienen un argumento `null_if_invalid` opcional. | 5 de octubre de 2017 |
| LISTAGG DISTINCT | Puede usar la cláusula DISTINCT con la función de agregación LISTAGG y la función de ventana LISTAGG para eliminar valores duplicados de la expresión especificada antes de la concatenación. | 5 de octubre de 2017 |
| Ver nombres de columna en mayúsculas | Para ver nombres de columna en resultados SELECT en mayúsculas, puede establecer el parámetro de configuración describe_field_name_in_uppercase en `true`. | 5 de octubre de 2017 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Omitir líneas de encabezado en tablas externas | Puede establecer la propiedad `skip.header.line.count` en el comando CREATE EXTERNAL TABLE para omitir las líneas de encabezado situadas al principio de los archivos de datos de Redshift Spectrum. | 5 de octubre de 2017 |
| Scan row count | Las reglas de monitorización de consultas de WLM usan la métrica scan_row_count para devolver el número de filas en un paso del análisis. La cantidad de filas es el número total de filas emitidas antes de filtrar las filas marcadas para eliminación (filas fantasma) y antes de aplicar los filtros de consulta definidos por el usuario. Para obtener más información, consulte Métricas de supervisión de consultas para Amazon Redshift aprovisionado. | 21 de septiembre de 2017 |
| Funciones de SQL definidas por el usuario | Una función SQL definida por el usuario (UDF) escalar incorpora una cláusula SQL SELECT que se ejecuta cuando se llama a la función y devuelve un único valor. Para obtener más información, consulte UDF de SQL escalares. | 31 de agosto de 2017 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Vistas de enlace en tiempo de ejecución | Una vista de enlace en tiempo de ejecución no está vinculada a los objetos de base de datos subyacentes, como tablas y funciones definidas por el usuario. Por tanto, no existe ninguna dependencia entre la vista y los objetos a los que hace referencia. Puede crear una vista aunque no existan los objetos a los que se hace referencia. Como no hay dependencias, puede quitar o modificar un objeto al que se referencie sin que esto afecte a la vista. Amazon Redshift no verifica si existen dependencias hasta que se consulta la vista. Para crear una vista de enlace en tiempo de ejecución, especifique la cláusula WITH NO SCHEMA BINDING en la instrucción CREATE VIEW. Para obtener más información, consulte CREATE VIEW. | 31 de agosto de 2017 |
| Función OCTET_LENGTH | OCTET_LENGTH devuelve la longitud de la cadena especificada como el número de bytes. | 18 de agosto de 2017 |
| Se admiten los tipos de archivos ORC y Grok | Ahora, Amazon Redshift Spectrum admite los formatos de datos ORC y Grok para los archivos de datos de Redshift Spectrum. Para obtener más información, consulte Archivos de datos para consultas en Amazon Redshift Spectrum. | 18 de agosto de 2017 |
| RegexSerDe ya es compatible | Amazon Redshift Spectrum ahora es compatible con el formato de datos RegexSerDe. Para obtener más información, consulte Archivos de datos para consultas en Amazon Redshift Spectrum. | 19 de julio de 2017 |
| Nuevas columnas añadidas a SVV_TABLES y SVV_COLUMNS | Se añadieron las columnas `domain_name` y `remarks` a SVV_COLUMNS. Se añadió una columna de comentarios a SVV_TABLES. | 19 de julio de 2017 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Vistas del sistema SVV_TABLES y SVV_COLUMNS | Las vistas del sistema SVV_TABLES y SVV_COLUMNS brindan información acerca de las columnas y otros detalles para vistas y tablas locales y externas. | 7 de julio de 2017 |
| Ya no se requiere una VPC para Amazon Redshift Sp con un metastore de Hive de Amazon EMR | Redshift Spectrum eliminó el requisito que indicaba que el clúster de Amazon Redshift y el clúster de Amazon EMR debían tener la misma VPC y la misma subred cuando utilizaban el metastore de Hive de Amazon EMR. Para obtener más información, consulte Uso de catálogos externos en Amazon Redshift Spectrum. | 7 de julio de 2017 |
| Utilizar el comando UNLOAD en tamaños de archivos más pequeños | De manera predeterminada, UNLOAD crea varios archivos en Amazon S3 con un tamaño máximo de 6,2 GB. Para crear archivos más pequeños, especifiq ue MAXFILESIZE con el comando UNLOAD. Puede especificar un tamaño de archivo máximo entre 5 MB y 6,2 GB. Para obtener más información, consulte UNLOAD. | 7 de julio de 2017 |
| TABLE PROPERTIES | Ahora puede establecer el parámetro numRows de TABLE PROPERTIES para CREATE EXTERNAL TABLE o ALTER TABLE con el fin de actualizar las estadísticas de la tabla para reflejar la cantidad de filas en la tabla. | 6 de junio de 2017 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| ANALYZE PREDICATE COLUMNS | Para ahorrar tiempo y recursos del clúster, puede optar por analizar solo las columnas que probablem ente utilice como predicados. Cuando ejecuta ANALYZE con la cláusula PREDICATE COLUMNS, la operación de análisis incluye solo las columnas que han sido utilizadas en una combinación, condición de filtro o cláusula de agrupación, o que se utilizan como clave de ordenación o clave de distribución. Para obtener más información, consulte Análisis de tablas. | 25 de mayo de 2017 |
| Políticas de IAM para Amazon Redshift Sp | Para conceder acceso a un bucket de Amazon S3 utilizando solo Redshift Spectrum, puede incluir una condición que le otorgue acceso al agente de usuario `AWS Redshift/Spectrum` . Para obtener más información, consulte Políticas de IAM para Amazon Redshift Spectrum. | 25 de mayo de 2017 |
| Análisis recursivo de Amazon Redshift Sp | Ahora, Redshift Spectrum analiza archivos en subcarpetas, así como en la carpeta especificada en Amazon S3. Para obtener más información, consulte Tablas externas para Redshift Spectrum. | 25 de mayo de 2017 |
| Reglas de monitorización de consultas | A través de las reglas de monitoreo de consultas de la WLM, puede definir los límites de rendimiento basados en métricas para las colas de WLM y especificar la acción que se debe realizar cuando una consulta supera esos límites: registro, salto y anulación. Puede definir las reglas de monitorización de consultas como parte de la configuración de la administración de cargas de trabajo (WLM). Para obtener más informaci ón, consulte Reglas de monitoreo de consultas de WLM. | 21 de abril de 2017 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Amazon Redshift Spectrum | Con Redshift Spectrum, puede consultar y recuperar datos de archivos en Amazon S3 de manera eficiente sin tener que cargar los datos en las tablas. Las consultas de Redshift Spectrum se ejecutan muy rápido en grandes conjuntos de datos, ya que Redshift Spectrum analiza los archivos de datos directamente en Amazon S3. Gran parte del procesamiento se produce en la capa de Amazon Redshift Spectrum, y la mayoría de los datos permanece en Amazon S3. Varios clústeres pueden consultar el mismo conjunto de datos en Amazon S3 de manera simultánea sin tener que generar copias de los datos para cada clúster. Para obtener más información, consultar Amazon Redshift Spectrum | 19 de abril de 2017 |
| Nuevas tablas del sistema serán compatibles con Redshift Spectrum | Se han añadido las siguientes vistas del sistema nuevas para admitir Redshift Spectrum:<br>• SVL_S3QUERY<br>• SVL_S3QUERY_SUMMARY<br>• SVV_EXTERNAL_COLUMNS<br>• SVV_EXTERNAL_DATABASES<br>• SVV_EXTERNAL_PARTITIONS<br>• SVV_EXTERNAL_TABLES<br>• PG_EXTERNAL_SCHEMA | 19 de abril de 2017 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Función de agregación APPROXIMATE PERCENTILE_DISC | La función de agregación APPROXIMATE PERCENTILE_DISC está disponible ahora. | 4 de abril de 2017 |
| Cifrado del lado del servidor con KMS | Ahora, puede descargar datos en Amazon S3 a través del cifrado en el lado del servidor con una clave de AWS Key Management Service (SSE-KMS). Además, COPY ahora carga archivos de datos cifrados por KMS de manera transparente desde Amazon S3. Para obtener más información, consulte UNLOAD. | 9 de febrero de 2017 |
| Nueva sintaxis de autorización | Ahora puede usar los parámetros IAM_ROLE, MASTER_SYMMETRIC_KEY, ACCESS_KEY_ID, SECRET_ACCESS_KEY y SESSION_TOKEN para proporcionar información de autorización y acceso para los comandos COPY, UNLOAD y CREATE LIBRARY. La nueva sintaxis de autorización proporcio na una alternativa más flexible a brindar un único argumento de cadena al parámetro CREDENTIALS. Para obtener más información, consulte Parámetros de autorización. | 9 de febrero de 2017 |
| Aumento del límite de esquemas | Ahora puede crear hasta 9900 esquemas por clúster. Para obtener más información, consulte CREATE SCHEMA. | 9 de febrero de 2017 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Codificación de tabla predeterminada | CREATE TABLE y ALTER TABLE ahora asignan codificación de compresión LZO a la mayoría de columnas nuevas. Por defecto, se asigna una codificación RAW a las columnas definidas como claves de ordenación, columnas que están definidas como tipos de datos BOOLEAN, REAL o DOUBLE PRECISION y tablas temporales. Para obtener más información, consulte ENCODE. | 6 de febrero de 2017 |
| Codificación de compresión ZSTD | Ahora, Amazon Redshift es compatible con la codificación de compresión de columnas ZSTD. | 19 de enero de 2017 |
| Funciones de agregación PERCENTILE_CONT y MEDIAN | PERCENTILE_CONT y MEDIAN ahora están disponibles como funciones de agregación y como funciones de ventana. | 19 de enero de 2017 |
| Registro del usuario de funciones definidas por el usuario (UDF) | Puede usar el módulo de registro de Python para crear mensajes de error y advertencia definidos por el usuario en sus User-Defined Scalar Function (UDF, Función escalar definida por el usuario). Después de la ejecución de una consulta, puede consultar la vista de sistema SVL_UDF_LOG para recuperar mensajes registrados. Para obtener más información acerca de los mensajes definidos por el usuario, consulte Registro de errores y advertencias en las UDF de Python. | 8 de diciembre de 2016 |
| Reducción estimada de ANALYZE COMPRESSION | El comando ANALYZE COMPRESSION ahora informa un cálculo del porcentaje de la reducción de espacio en disco para cada columna. Para obtener más información, consulte ANALYZE COMPRESSION. | 10 de noviembre de 2016 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Límites de conexión | Ahora puede establecer un límite sobre la cantidad de conexiones a la base de datos que un usuario puede tener abiertas al mismo tiempo. También puede limitar la cantidad de conexiones simultáneas para una base de datos. Para obtener más información, consulte CREAR USUARIO y CREATE DATABASE. | 10 de noviembre de 2016 |
| Mejoras en el orden de COPY | COPY ahora añade automáticamente nuevas filas a la región ordenada de la tabla cuando carga los datos en el orden de la clave de ordenación. Para ver los requisitos específicos para habilitar esta mejora, consulte Carga de los datos en orden de clave de clasificación | 10 de noviembre de 2016 |
| CTAS con compresión | CREATE TABLE AS (CTAS) ahora asigna automátic amente codificaciones de compresión a tablas nuevas en función del tipo de datos de la columna. Para obtener más información, consulte Herencia de atributos de columna y tabla. | 28 de octubre de 2016 |
| Marca temporal con tipo de datos de zona horaria | Ahora, Amazon Redshift admite una marca temporal con el tipo de datos de zona horaria (TIMESTAMPTZ). También, se añadieron nuevas funciones para admitir el nuevo tipo de datos. Para obtener más información, consulte Funciones de fecha y hora. | 29 de septiembre de 2016 |
| Umbral de análisis | Para reducir el tiempo de procesamiento y mejorar el rendimiento de las operaciones ANALYZE, Amazon Redshift omite el análisis de una tabla si el porcentaje de filas que ha cambiado desde que se ejecutó el último comando ANALYZE es inferior al umbral de análisis especificado por el parámetro analyze_threshold_percent. Por defecto, `analyze_threshold_percent` es 10. | 9 de agosto de 2016 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Nueva tabla del sistema STL_RESTARTED_SESSIONS | Cuando Amazon Redshift reinicia una sesión, STL_RESTARTED_SESSIONS registra el nuevo ID de proceso (PID) y el PID anterior. | 9 de agosto de 2016 |
| Actualización de la documentación sobre funciones de fecha y hora | Se añadió un resumen de las funciones con enlaces al Funciones de fecha y hora, y se actualizaron las referencias de las funciones para mantener la consistencia. | 24 de junio de 2016 |
| Nuevas columnas en STL_CONNECTION_LOG | La tabla del sistema STL_CONNECTION_LOG tiene dos columnas nuevas para realizar un seguimiento de las conexiones SSL. Si habitualmente carga registros de auditoría a una tabla de Amazon Redshift, necesitará añadir las siguientes columnas nuevas a la tabla de destino: sslcompression y sslexpansion. | 5 de mayo de 2016 |
| Contraseña hash de MD5 | Puede especificar la contraseña para un comando CREAR USUARIO o ALTER USER al proporcionar la cadena hash de MD5 de la contraseña y el nombre de usuario. | 21 de abril de 2016 |
| Nueva columna en STV_TBL_PERM | La columna `backup` en la vista del sistema STV_TBL_PERM indica si la tabla se incluye en las instantáneas del clúster. Para obtener más información, consulte BACKUP. | 21 de abril de 2016 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Tablas sin backup | Para las tablas, como las tablas provisionales que no contienen datos críticos, puede especificar BACKUP NO en la instrucción CREATE TABLE o CREATE TABLE AS para evitar que Amazon Redshift incluya la tabla en instantáneas automatizadas o manuales. Si utiliza tablas sin copia de seguridad, se guarda la hora de procesamiento cuando crea instantáneas y restaura los datos a partir de las instantáneas. Además, reduce el espacio de almacenamiento en Amazon S3. | 7 de abril de 2016 |
| Umbral de eliminación de VACUUM | Por defecto, el comando VACUUM recupera espacio para que, al menos, el 95 por ciento de las filas restantes no se marque para la eliminación. Como resultado, VACUUM normalmente necesita mucho menos tiempo para la fase de eliminación comparado con la recuperación del 100 por ciento de filas eliminadas. Para cambiar el umbral predeterminado de una tabla, incluya el parámetro TO threshold PERCENT cuando ejecute el comando VACUUM. | 7 de abril de 2016 |
| Tabla del sistema SVV_TRANS ACTIONS | La vista del sistema SVV_TRANSACTIONS registra información acerca de transacciones que tienen bloqueos actualmente en las tablas de la base de datos. | 7 de abril de 2016 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Uso de roles IAM para obtener acceso a otros recursos de AWS | Para mover datos entre el clúster y otro recurso de AWS, como Amazon S3, DynamoDB, Amazon EMR o Amazon EC2, el clúster debe tener permiso para obtener acceso al recurso y realizar las acciones necesarias. Como una alternativa más segura de proporcionar un par de claves de acceso con comandos COPY, UNLOAD o CREATE LIBRARY, ahora puede especificar un rol IAM que el clúster utiliza para autenticación y autorización. Para obtener más información, consulte Control de acceso con base en roles. | 29 de marzo de 2016 |
| Umbral de ordenación de VACUUM | El comando VACUUM ahora omite la fase de ordenación para cualquier tabla que tenga más del 95 por ciento de las filas de la tabla ordenadas. Para cambiar el umbral de ordenación predeterminado de una tabla, incluya el parámetro TO threshold PERCENT cuando ejecute el comando VACUUM. | 17 de marzo de 2016 |
| Nuevas columnas en STL_CONNECTION_LOG | La tabla del sistema STL_CONNECTION_LOG tiene tres columnas nuevas. Si habitualmente carga registros de auditoría a una tabla de Amazon Redshift, tendrá que agregar las siguientes columnas nuevas a la tabla de destino: sslversion, sslcipher y mtu. | 17 de marzo de 2016 |
| Uso del comando UNLOAD con compresión bzip2 | Ahora tiene la opción de UNLOAD a través de la compresión bzip2. | 8 de febrero de 2016 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| ALTER TABLE APPEND | ALTER TABLE APPEND adjunta filas en una tabla de destino al mover datos de una tabla de origen existente. Por lo general, ALTER TABLE APPEND es mucho más rápida que una operación CREATE TABLE AS o INSERT INTO similar debido a que los datos se mueven, no se duplican. | 8 de febrero de 2016 |
| Salto de cola de consultas de WLM | Si la administración de cargas de trabajo (WLM) cancela una consulta de solo lectura, como una instrucción SELECT, debido a un tiempo de espera de WLM, WLM intentará dirigir la consulta a la siguiente cola coincidente. Para obtener más información, consultar Salto de cola de consultas de WLM | 7 de enero de 2016 |
| ALTER DEFAULT PRIVILEGES | Puede usar el comando ALTER DEFAULT PRIVILEGES para definir el conjunto predeterminado de privilegios de acceso que se aplicará a los objetos que el usuario especificado crea en el futuro. | 10 de diciembre de 2015 |
| Compresión de archivos bzip2 | El comando COPY admite la carga de datos de archivos que se comprimieron con bzip2. | 10 de diciembre de 2015 |
| NULLS FIRST y NULLS LAST | Puede especificar si una cláusula ORDER BY debe clasificar los valores NULLS en primer o último lugar en el conjunto de resultados. Para obtener más información, consulte Cláusula ORDER BY y Resumen de la sintaxis de la función de ventana. | 19 de noviembre de 2015 |
| Palabra clave REGION para CREATE LIBRARY | Si el bucket de Amazon S3 que contiene los archivos de la biblioteca de UDF no se encuentra en la misma región de AWS que el clúster de Amazon Redshift, puede usar la opción REGION para especificar la región en la que se ubican los datos. Para obtener más información, consulte CREATE LIBRARY. | 19 de noviembre de 2015 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Funciones escalares definidas por el usuario (UDF) | Ahora, puede crear funciones personalizadas escalares definidas por el usuario para implement ar una funcionalidad de procesamiento que no sea de SQL proporcionada por los módulos admitidos por Amazon Redshift en la Biblioteca estándar de Python 2.7 o en sus propias UDF basadas en el lenguaje de programación Python. Para obtener más información, consulte Funciones definidas por el usuario en Amazon Redshift. | 11 de septiembre de 2015 |
| Propiedades dinámicas en la configuración de WLM | El parámetro de configuración de WLM ahora admite la aplicación de algunas propiedades de manera dinámica. Otras propiedades son cambios estáticos y requieren que los clústeres asociados se reinicien para que se apliquen los cambios de la configuración. Para obtener más información, consulte Propiedad es de configuración dinámicas y estáticas de WLM y Administración de la carga de trabajo. | 3 de agosto de 2015 |
| Función LISTAGG | Función LISTAGG y Función de ventana LISTAGG devuelven una cadena creada por la concatenación de un conjunto de valores de columna. | 30 de julio de 2015 |
| Parámetro no disponible | El parámetro de configuración max_cursor_result_ set_size no está disponible. El tamaño de los conjuntos de resultados del cursor se basa en el tipo de nodo del clúster. Para obtener más información, consulte Restricciones del cursor. | 24 de julio de 2015 |
| Documentación revisada sobre el comando COPY | Se han revisado las referencias del comando COPY para que el material sea más simple y se pueda obtener acceso a él fácilmente. | 15 de julio de 2015 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Uso del comando COPY del formato Avro | El comando COPY admite la carga de datos en formato Avro de archivos de datos en Amazon S3, Amazon EMR y alojamiento remoto a través de SSH. Para obtener más información, consulte AVRO y Ejemplos de Copy de Avro. | 8 de julio de 2015 |
| STV_STARTUP_RECOVERY_STATE | La tabla de sistema STV_STARTUP_RECOVERY_STATE registra el estado de las tablas que están bloqueadas temporalmente durante las operaciones de reinicio del clúster. Amazon Redshift bloquea de manera temporal las tablas mientras estas se procesan para resolver transacciones obsoletas después de que se reinicia un clúster. | 25 de mayo de 2015 |
| ORDER BY opcional para las funciones de clasificación | La cláusula ORDER BY ahora es opcional para determinadas funciones de clasificación de ventana. Para obtener más información, consulte Función de ventana CUME_DIST, Función de ventana DENSE_RANK, Función de ventana RANK, Función de ventana NTILE, Función de ventana PERCENT_RANK y Función de ventana ROW_NUMBER. | 25 de mayo de 2015 |
| Claves de ordenación intercaladas | Las claves de ordenación intercaladas le otorgan el mismo peso a cada columna en la clave de ordenación. Al utilizar claves de ordenación intercaladas en lugar de claves compuestas predeterminadas, se mejora el rendimiento considerablemente para las consultas que usan predicados restrictivos en las columnas de ordenación secundarias, especialmente para tablas grandes. La ordenación intercalada también mejora el rendimiento general cuando varias consultas se filtran en diferentes columnas de la misma tabla. Para obtener más información, consulte Claves de clasificación y CREATE TABLE. | 11 de mayo de 2015 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Tema revisado sobre el ajuste del rendimiento de las consultas | Se ha ampliado Ajuste del rendimiento de las consultas para incluir nuevas consultas para el análisis de rendimiento de las consultas y más ejemplos. Además, el tema se ha revisado para que sea más claro y completo. Prácticas recomendadas de Amazon Redshift para el diseño de consultas contiene más información acerca de cómo escribir consultas para mejorar el rendimiento. | 23 de marzo de 2015 |
| SVL_QUERY_QUEUE_INFO | La vista SVL_QUERY_QUEUE_INFO resume los detalles de las consultas que pasan tiempo en una cola de consultas o en una cola de confirmación de WLM. | 19 de febrero de 2015 |
| SVV_TABLE_INFO | Puede utilizar la vista SVV_TABLE_INFO para diagnosticar y abordar problemas en el diseño de la tabla que pueden influir en el rendimiento de las consultas, incluidos problemas con codificación de compresión, claves de distribución, estilo de ordenación, sesgo de distribución de datos, tamaño de tabla y estadísticas. | 19 de febrero de 2015 |
| El comando UNLOAD utiliza el cifrado de archivos del lado del servidor | Ahora, el comando UNLOAD utiliza automáticamente cifrado del lado del servidor (SSE) de Amazon S3 para cifrar todos los archivos de datos de descarga. El cifrado del lado del servidor añade otra capa de seguridad de datos con ningún cambio, o pocos, en el rendimiento. | 31 de octubre de 2014 |
| Función de ventana CUME_DIST | La Función de ventana CUME_DIST calcula la distribución acumulada de un valor dentro de una ventana o partición. | 31 de octubre de 2014 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Función MONTHS_BETWEEN | La Función MONTHS_BETWEEN determina la cantidad de meses que hay entre las dos fechas. | 31 de octubre de 2014 |
| Función NEXT_DAY | La Función NEXT_DAY devuelve la fecha de la primera instancia de un día especificado que es posterior a la fecha proporcionada. | 31 de octubre de 2014 |
| Función de ventana PERCENT_RANK | La Función de ventana PERCENT_RANK calcula la clasificación de porcentaje de una fila dada. | 31 de octubre de 2014 |
| Función de ventana RATIO_TO_REPORT | La Función de ventana RATIO_TO_REPORT calcula la relación de un valor con la suma de los valores en una ventana o partición. | 31 de octubre de 2014 |
| Función TRANSLATE | La Función TRANSLATE reemplaza todas las coincidencias de caracteres especificados dentro de una expresión dada con sustitutos especificados. | 31 de octubre de 2014 |
| Función NVL2 | La Función NVL2 devuelve uno de los dos valores según si una expresión especificada toma un valor NULL o NOT NULL. | 16 de octubre de 2014 |
| Función de ventana MEDIAN | La Función de ventana MEDIAN calcula el valor mediano para el rango de valores en una ventana o partición. | 16 de octubre de 2014 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Cláusula ON ALL TABLES IN SCHEMA schema_name para los comandos GRANT y REVOKE | Se actualizaron los comandos GRANT y REVOKE con una cláusula ON ALL TABLES IN SCHEMA schema_name. Esta cláusula le permite usar un comando para cambiar los privilegios de todas las tablas de un esquema. | 16 de octubre de 2014 |
| Cláusula IF EXISTS para los comandos DROP SCHEMA, DROP TABLE, DROP USER y DROP VIEW | Se actualizaron los comandos DROP SCHEMA, DROP TABLE, DROP USER y DROP VIEW con una cláusula IF EXISTS. Esta cláusula hace que el comando no realice cambios y devuelva un mensaje en lugar de terminar con un error si el objeto especificado no existe. | 16 de octubre de 2014 |
| Cláusula IF NOT EXISTS para los comandos CREATE SCHEMA y CREATE TABLE | Se actualizaron los comandos CREATE SCHEMA y CREATE TABLE con una cláusula IF NOT EXISTS. Esta cláusula hace que el comando no realice cambios y devuelva un mensaje en lugar de terminar con un error si el objeto especificado ya existe. | 16 de octubre de 2014 |
| Compatibilidad de COPY con la codificación UTF-16 | El comando COPY ahora admite la carga de archivos de datos que utilizan codificación UTF-16 y codificación UTF-8. Para obtener más información, consulte ENCODING. | 29 de septiembre de 2014 |
| Nuevo tutorial de administración de cargas de trabajo | Tutorial: Configuración de colas de Administración de carga de trabajo (WLM) manual le enseña el proceso de configuración de colas de administración de cargas de trabajo (WLM) para mejorar el procesamiento de consultas y la asignación de recursos. | 25 de septiembre de 2014 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Cifrado AES de 128 bits | Ahora, el comando COPY admite el cifrado AES de 128 bits y el cifrado AES de 256 bits cuando carga datos desde archivos de datos cifrados a través del cifrado del lado del cliente de Amazon S3. Para obtener más información, consulte Carga de archivos de datos cifrados desde Amazon S3. | 29 de septiembre de 2014 |
| Función PG_LAST_UNLOAD_COUNT | La función PG_LAST_UNLOAD_COUNT devuelve la cantidad de filas que se procesaron en la operación UNLOAD más reciente. Para obtener más información, consulte PG_LAST_UNLOAD_COUNT. | 15 de septiembre de 2014 |
| Nueva sección de solución de problemas de consultas | Solución de problemas de consultas proporciona una referencia rápida para identificar y abordar algunos de los problemas más frecuentes y más graves que posiblemente enfrente con las consultas de Amazon Redshift. | 7 de julio de 2014 |
| Nuevo tutorial de carga de datos | En Tutorial: Carga de datos desde Amazon S3, descubrirá el proceso para cargar datos en las tablas de bases de datos de Amazon Redshift desde archivos de datos en un bucket de Amazon S3 de principio a fin. | 1 de julio de 2014 |
| Función de ventana PERCENTILE_CONT | Función de ventana PERCENTILE_CONT es una función de distribución inversa que asume un modelo de distribución continua. Toma un valor percentil y una especificación de ordenación, y devuelve un valor interpolado que se encuadraría dentro de ese valor percentil dado respecto de la especificación de ordenación. | 30 de junio de 2014 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Función de ventana PERCENTILE_DISC | Función de ventana PERCENTILE_DISC es una función de distribución inversa que asume un modelo de distribución discreta. Toma un valor percentil y una especificación de orden, y devuelve un elemento del conjunto. | 30 de junio de 2014 |
| Funciones GREATEST y LEAST | Las funciones Funciones GREATEST y LEAST devuelven el valor más grande o más pequeño de una lista de expresiones. | 30 de junio de 2014 |
| Comando COPY entre regiones | El comando COPY admite la carga de datos de un bucket de Amazon S3 o una tabla de Amazon DynamoDB que estén ubicados en una región diferente a la del clúster de Amazon Redshift. Para obtener más información, consulte REGION en COPY command reference. | 30 de junio de 2014 |
| Prácticas recomendadas ampliadas | Se ha ampliado, reorganizado y movido Prácticas recomendadas de Amazon Redshift a la parte superior de la jerarquía de navegación para que se pueda detectar con mayor facilidad. | 28 de mayo de 2014 |
| Utilizar el comando UNLOAD en un único archivo | El comando UNLOAD puede, de manera opcional, descargar datos de tablas en serie en un único archivo de Amazon S3 al agregar la opción PARALLEL OFF. Si el tamaño de los datos es superior al tamaño máximo de archivo de 6,2 GB, UNLOAD crea archivos adicionales. | 6 de mayo de 2014 |
| Funciones REGEXP | Las funciones REGEXP_COUNT, REGEXP_INSTR y REGEXP_REPLACE manipulan cadenas basadas en el patrón de expresión regular coincidente. | 6 de mayo de 2014 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| COPY de Amazon EMR | El comando COPY admite la carga de datos directamente desde los clústeres de Amazon EMR. Para obtener más información, consulte Carga de datos desde Amazon EMR. | 18 de abril de 2014 |
| Aumento del límite de simultaneidad de WLM | Ahora puede configurar la administración de cargas de trabajo (WLM) para que ejecute hasta 50 consultas simultáneamente en las colas de consultas definidas por el usuario. Este aumento les proporciona a los usuarios mayor flexibilidad para administrar el rendimiento del sistema al modificar las configuraciones de WLM. Para obtener más información, consultar Implementación de WLM manual | 18 de abril de 2014 |
| Nuevo parámetro de configuración para administrar el tamaño del cursor | El parámetro de configuración `max_cursor_result_set_size` define el tamaño máximo de datos, en megabytes, que pueden devolverse por conjunto de resultados del cursor en una consulta más grande. El valor de este parámetro también afecta la cantidad de cursores simultáneos para el clúster, lo que le permite configurar un valor que aumente o disminuya la cantidad de cursores del clúster.<br><br>Para obtener más información, consulte DECLARE en esta guía y Configuración del tamaño máximo del conjunto de resultados de un cursor en la Guía de administración de Amazon Redshift. | 28 de marzo de 2014 |
| COPY de formato JSON | El comando COPY admite la carga de datos en formato JSON de archivos de datos en Amazon S3 y de alojamiento remoto a través de SSH. Para obtener más información, consulte las notas de uso COPY de formato JSON. | 25 de marzo de 2014 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Nueva tabla del sistema STL_PLAN_INFO | La tabla STL_PLAN_INFO complementa el comando EXPLAIN como otra manera de observar planes de consultas. | 25 de marzo de 2014 |
| Nueva función REGEXP_SUBSTR | La Función REGEXP_SUBSTR devuelve los caracteres extraídos de una cadena al buscar un patrón de expresión regular. | 25 de marzo de 2014 |
| Nuevas columnas para STL_COMMIT_STATS | La tabla STL_COMMIT_STATS tiene dos columnas nuevas: `numxids` y `oldestxid` . | 6 de marzo de 2014 |
| Compatibilidad de COPY con SSH para gzip y lzop | El comando COPY admite la compresión de gzip y lzop cuando carga datos a través de una conexión SSH. | 13 de febrero de 2014 |
| Nuevas funciones | La Función de ventana ROW_NUMBER devuelve el número de la fila actual. Función STRTOL convierte una expresión de cadena de un número de la base especificada al valor entero equivalente. PG_CANCEL_BACKEND y PG_TERMINATE_BACKEND permiten a los usuarios cancelar consultas y conexiones de sesiones. Se ha añadido la función LAST_DAY para la compatibilidad con Oracle. | 13 de febrero de 2014 |
| Nueva tabla del sistema | La tabla del sistema STL_COMMIT_STATS proporciona métricas relacionadas con el rendimiento de las confirmaciones, incluido el tiempo de varias fases de confirmación y el número de bloques confirmados. | 13 de febrero de 2014 |
| FETCH con clústeres de un nodo | Cuando utiliza un cursor en un clúster de un nodo, la cantidad máxima de filas que puede recuperar a través del comando FETCH es 1 000. FETCH FORWARD ALL no es compatible con los clústeres de un nodo. | 13 de febrero de 2014 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Estrategia de redistribución DS_DIST_ALL_INNER | DS_DIST_ALL_INNER en el resultado del plan de explicación indica que una tabla interna completa se redistribuyó a un único sector debido a que la tabla externa utiliza DISTSTYLE ALL. Para obtener más información, consulte Ejemplos de tipos de combinación y Evaluación del plan de consulta. | 13 de enero de 2014 |
| Nuevas tablas del sistema para consultas | Amazon Redshift ha agregado nuevas tablas de sistema que los clientes pueden utilizar para evaluar la ejecución de consultas destinadas al ajuste y la resolución de problemas. Para obtener más información, consulte SVL_COMPILE, STL_SCAN, STL_RETURN, STL_SAVE STL_ALERT_EVENT_LOG. | 13 de enero de 2014 |
| Cursores de un nodo | Los cursores ahora son compatibles con los clústeres de un nodo. Un clúster de un nodo puede tener dos cursores al mismo tiempo, con un conjunto de resultados máximo de 32 GB. En un clúster de un nodo, recomendamos ajustar el parámetro del tamaño de caché ODBC en 1 000. Para obtener más información, consulte DECLARE. | 13 de diciembre de 2013 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Estilo de distribución ALL | La distribución ALL puede acortar considerablemente los tiempos de ejecución para determinados tipos de consultas. Cuando una tabla utiliza el estilo de distribución ALL, una copia de toda la tabla se distribuye a cada nodo. Debido a que la tabla se coloca efectivamente con todas las otras tablas, no se necesita una redistribución durante la ejecución de la consulta. La distribución ALL no es adecuada para todas las tablas ya que aumenta los requisitos de almacenamiento y el tiempo de carga. Para obtener más información, consulte Distribución de datos para la optimización de consultas. | 11 de noviembre de 2013 |
| COPY de hosts remotos | Además de cargar tablas de archivos de datos en Amazon S3 y de tablas de Amazon DynamoDB, el comando COPY puede cargar datos de texto desde clústeres de Amazon EMR, instancias de Amazon EC2 y otros alojamientos remotos a través de conexiones SSH. Amazon Redshift utiliza varias conexiones SSH simultáneas para leer y cargar datos en paralelo. Para obtener más información, consulte Carga de datos desde hosts remotos. | 11 de noviembre de 2013 |
| Porcentaje de memoria de WLM utilizada | Puede equilibrar la carga de trabajo al designar un porcentaje específico de memoria para cada cola en la configuración de la administración de cargas de trabajo (WLM). Para obtener más información, consulte Implementación de WLM manual. | 11 de noviembre de 2013 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| APPROXIMATE COUNT(DISTINCT) | Las consultas que utilizan APPROXIMATE COUNT(DISTINCT) se ejecutan mucho más rápido, con un error relativo de alrededor del 2%. La función APPROXIMATE COUNT(DISTINCT) utiliza un algoritmo HyperLogLog. Para obtener más información, consulte Función COUNT. | 11 de noviembre de 2013 |
| Nuevas funciones SQL para recuperar detalles de consultas recientes | Cuatro nuevas funciones SQL recuperan detalles acerca de consultas y comandos COPY recientes. Las nuevas funciones facilitan la consulta de tablas de registro del sistema y, en mucho casos, brindan los detalles necesarios sin la necesidad de obtener acceso a las tablas del sistema. Para obtener más información, consulte PG_BACKEND_PID, PG_LAST_COPY_ID, PG_LAST_COPY_COUNT, PG_LAST_QUERY_ID. | 1 de noviembre de 2013 |
| Opción MANIFEST para UNLOAD | La opción MANIFEST para el comando UNLOAD complementa la opción MANIFEST para el comando COPY. Si utiliza la opción MANIFEST con UNLOAD, se crea automáticamente un archivo de manifiesto que enumera de forma explícita los archivos de datos que se crearon en Amazon S3 con la operación de descarga. Luego, puede utilizar el mismo archivo de manifiesto con un comando COPY para cargar los datos. Para obtener más información, consulte Descarga de datos en Amazon S3 y Ejemplos de UNLOAD. | 1 de noviembre de 2013 |
| Opción MANIFEST para COPY | Puede utilizar la opción MANIFEST con el comando COPY para enumerar de manera explícita los archivos de datos que se cargarán desde Amazon S3. | 18 de octubre de 2013 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Tablas del sistema para la solución de problemas de consultas | Se añadió documentación para las tablas del sistema que se utilizan para solucionar problemas en consultas . La sección Vistas STL para registro ahora contiene documentos para las siguientes tablas del sistema: STL_AGGR, STL_BCAST, STL_DIST, STL_DELETE, STL_HASH, STL_HASHJOIN, STL_INSERT, STL_LIMIT, STL_MERGE, STL_MERGEJOIN, STL_NESTLOOP, STL_PARSE, STL_PROJECT, STL_SCAN, STL_SORT, STL_UNIQUE, STL_WINDOW. | 3 de octubre de 2013 |
| Función CONVERT_TIMEZONE | La Función CONVERT_TIMEZONE convierte una marca temporal de una zona horaria a otra, con la opción de ajustar automáticamente el horario de verano. | 3 de octubre de 2013 |
| Función SPLIT_PART | La Función SPLIT_PART divide una cadena en el delimitador especificado y devuelve la parte en la posición especificada. | 3 de octubre de 2013 |
| Tabla del sistema STL_USERLOG | STL_USERLOG registra detalles de cambios que ocurren cuando se crea, modifica o elimina un usuario de la base de datos. | 3 de octubre de 2013 |
| Codificación de columnas LZO y compresión de archivos LZOP. | La codificación de compresión de columnas LZO combina una relación de compresión muy alta con un buen rendimiento. COPY de Amazon S3 admite la carga de archivos comprimidos a través de la compresión LZOP. | 19 de septiembre de 2013 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| JSON, expresiones regulares y cursores | Se añadió compatibilidad para el análisis de cadenas JSON, la coincidencia de patrones a través de expresiones regulares y el uso de cursores para recuperar grandes conjuntos de datos mediante una conexión ODBC. Para obtener más información, consulte Funciones JSON, Condiciones de coincidencia de patrones y DECLARE. | 10 de septiembre de 2013 |
| Opción ACCEPTINVCHAR para COPY | Puede cargar datos que contengan caracteres UTF-8 no válidos al especificar la opción ACCEPTINVCHAR con el comando COPY. | 29 de agosto de 2013 |
| Opción CSV para COPY | El comando COPY ahora admite la carga de archivos CSV de entrada formateados. | 9 de agosto de 2013 |
| CRC32 | Función CRC32 realiza comprobaciones de redundancia cíclicas. | 9 de agosto de 2013 |
| Comodines de WLM | La administración de cargas de trabajo (WLM) admiten comodines para añadir grupos de usuarios y grupos de consultas a colas. Para obtener más información, consulte Caracteres comodín. | 1 de agosto de 2013 |
| Tiempo de espera de WLM | Para limitar la cantidad de tiempo que las consultas de una determinada cola de WLM tienen permitido utilizar, puede establecer el valor de tiempo de espera de WLM para cada cola. Para obtener más información, consulte Tiempo de espera de WLM. | 1 de agosto de 2013 |
| Nuevas opciones COPY 'auto' y 'epochsecs' | El comando COPY realiza un reconocimiento automático de formatos de fecha y hora. Los nuevos formatos de tiempo, 'epochsecs' y 'epochmillisecs' permiten que COPY cargue datos en el formato de fecha de inicio. | 25 de julio de 2013 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Función CONVERT_TIMEZONE | La Función CONVERT_TIMEZONE convierte una marca temporal de una zona horaria a otra. | 25 de julio de 2013 |
| Función FUNC_SHA1 | La Función FUNC_SHA1 convierte una cadena a través del algoritmo SHA1. | 15 de julio de 2013 |
| max_execution_time | Para limitar la cantidad de tiempo que las consultas tienen permitido utilizar, puede establecer el parámetro max_execution_time como parte de la configuración de WLM. Para obtener más información, consulte Modificación de la configuración de WLM. | 22 de julio de 2013 |
| Caracteres UTF-8 de cuatro bytes | El tipo de datos VARCHAR ahora es compatible con caracteres UTF-8 de cuatro bytes. Los caracteres UTF-8 de cinco bytes o más no son compatibles. Para obtener más información, consulte Almacenamiento y rangos. | 18 de julio de 2013 |
| SVL_QERROR | La vista del sistema SVL_QERROR no está disponible. | 12 de julio de 2013 |
| Historial de revisión modificado | La página del historial de revisión ahora muestra la fecha en que se actualizó la documentación. | 12 de julio de 2013 |
| STL_UNLOAD_LOG | STL_UNLOAD_LOG registra detalles para una operación de descarga. | 5 de julio de 2013 |

| Cambio | Descripción | Fecha de modificación |
| --- | --- | --- |
| Parámetro de tamaño de búsqueda de JDBC | Para evitar errores de falta de memoria del lado del cliente cuando recupera grandes conjuntos de datos a través de JDBC, puede habilitar el cliente para recuperar datos en lotes al configurar el parámetro de tamaño de búsqueda de JDBC. Para obtener más información, consulte Configuración del parámetro de tamaño de búsqueda de la JDBC. | 27 de junio de 2013 |
| Comando UNLOAD para archivos cifrados | UNLOAD ahora admite la descarga de datos de una tabla en archivos cifrados en Amazon S3. | 22 de mayo de 2013 |
| Credenciales temporales | COPY y UNLOAD ahora admiten el uso de credencia les temporales. | 11 de abril de 2013 |
| Se añadieron clarificaciones | Se aclararon y ampliaron debates sobre el diseño de tablas y la carga de datos. | 14 de febrero de 2013 |
| Se agregaron prácticas recomendadas | Se añadieron Prácticas recomendadas de Amazon Redshift para el diseño de tablas y Prácticas recomendadas de Amazon Redshift para la carga de datos. | 14 de febrero de 2013 |
| Se aclararon las restricciones de contraseña | Se aclararon las restricciones de contraseña para CREATE USER y ALTER USER, y se añadieron revisiones menores. | 14 de febrero de 2013 |
| Nueva guía | Esta es la primera versión de la Guía para desarroll adores de Amazon Redshift. | 14 de febrero de 2013 |